Universidad Carlos III de Madrid
www.uc3m.es

# TESIS DOCTORAL

## GOODNESS-OF-FIT IN MULTIVARIATE TIME SERIES


**Autor:**

***Huong Nguyen Thu***

**Director:**

**Santiago Velilla Cerdán**

**DEPARTAMENTO DE ESTADÍSTICA**


Getafe, febrero 2014

Universidad Carlos III de Madrid
www.uc3m.es

## TESIS DOCTORAL

# GOODNESS-OF-FIT IN MULTIVARIATE TIME SERIES


***Autor:*** *Huong Nguyen Thu*

**Director:** **Santiago Velilla Cerdán**


Firma del Tribunal Calificador:

Firma

Presidente:

Vocal:

Secretario:

Calificación:

Getafe, de de

# Goodness-of-fit in Multivariate Time Series

A dissertation submitted for the degree of

*Doctor of Philosophy*

**Department of Statistics**

Universidad Carlos III de Madrid

Huong Nguyen Thu

Supervisor: Santiago Velilla

Universidad Carlos III de Madrid, 2013

To my family and professors.

# Acknowlegements

I would like to thank Dr.Prof Santiago Velilla for his sharing, encouragement and guidance as my supervisor. I am truly indebted and thankful to Esther Ruíz. I would also like to thank Antoni Espasa, Francisco Javier Prieto, Mike Wiper, Andrés M. Alonso, Helena Veiga, Agnieszka Jach and Isabel Molina. Besides, I am very grateful to Universidad Carlos III de Madrid, especially all members in Department of Statistics for all their support and interest. Lastly, and most importantly, I wish to thank my parents, my grandparents, many of my professors and friends for all the emotional support and caring they provided.

# Abstract

Goodness-of-fit is an important task in time series analysis. In this thesis, we propose a new family of statistics and a new goodness-of-fit process for the well-known multivariate autoregressive moving average *VARMA(p,q)* model.

Some preliminary results are studied first for an initial goodness-of-fit method. Since the residuals of the fit play an important role in identification and diagnostic checking, relations between least squares residuals and true errors are studied. An explicit representation of the information matrix as a limit is also obtained.

Second, we generalize a univariate goodness-of-fit process studied in Ubierna and Velilla (2007). An explicit form of the limit covariance function is presented, as well as a characterization of its limit properties in terms of a parametric Gaussian process. This motivates the introduction of a new goodness-of-fit process based on a transformed correlation matrix sequence. The construction and properties of the associated transformation matrices are investigated. We also prove the convergence of this new process to the Brownian bridge. Thus, statistics defined as functionals of our process use a null distribution that is free of unknown parameters.

Finally, simulations, comparisons, and examples of application are presented to illustrate our theoretical findings and contributions. Our proposed goodness-of-fit statistics are shown to be quite sensitive for detecting lack of fit. They also seem to be relatively independent of the choice of a particular lag.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

**Summary.** Diagnostic checks play an important role in many empirical studies. Techniques for checking the adequacy of univariate time series models have become widespread in both statistics and econometrics. Recently, there has been a great deal of interest in multivariate time series models. This chapter defines, in section 1.1, a commonly used model in this field, the multivariate autoregressive moving average *VARMA(p,q)* process. A review of literature for goodness-of-fit in univariate and multivariate time series is considered in sections 1.2 and 1.3, respectively. Section 1.4 introduces our research goals.

## 1.1 Model definition

The basic tool in this memory will be a causal and invertible $m-$variate autoregressive moving average *VARMA(p,q)* process of the form

$$\mathbf{\Phi}(B)(\mathbf{X}_t - \boldsymbol{\mu}) = \mathbf{\Theta}(B)\boldsymbol{\varepsilon}_t \ , \tag{1.1}$$

where $B$ is backward shift operator $B\mathbf{X}_t = \mathbf{X}_{t-1}$; $\boldsymbol{\mu}$ is the $m \times 1$ mean vector; and $\{\boldsymbol{\varepsilon}_t : t \in \mathbf{Z}\}$ is a zero mean white noise sequence $WN(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a $m \times m$ positive definite matrix. Additionally, $\mathbf{\Phi}(z) = \mathbf{I}_m - \mathbf{\Phi}_1 z - \cdots - \mathbf{\Phi}_p z^p$ and $\mathbf{\Theta}(z) = \mathbf{I}_m + \mathbf{\Theta}_1 z + \cdots + \mathbf{\Theta}_q z^q$ are matrix polynomials, where $\mathbf{I}_m$ is the $m \times m$ identity matrix, and $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p; \mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_q$ are $m \times m$ real matrices such that the roots of the determinantal equations $|\mathbf{\Phi}(z)| = 0$ and $|\mathbf{\Theta}(z)| = 0$ all lie outside the unit circle. It will be assumed that the $m(p+q)$ roots are different from each other. It will be also assumed that the conditions of Dunsmuir and Hannan (1976) hold, so that the model (1.1) is properly identified. Hence, the terms of the operator $\mathbf{\Phi}^{-1}(B)\mathbf{\Theta}(B) = \sum_{j=0}^{\infty} \mathbf{\Omega}_j B^j$ are uniquely defined. In particular, both $\mathbf{\Phi}_p$ and $\mathbf{\Theta}_q$ are non-null matrices, and $r(\mathbf{\Phi}_p, \mathbf{\Theta}_q) = m$ (Hannan, 1969).

In what follows, it will be convenient to put $P = \max(p,q)$, and to define the $m \times mp$ matrix $\mathbf{\Phi} = (\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p)$; the $m \times mq$ matrix $\mathbf{\Theta} = (\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_q)$; and the $m^2(p+q) \times 1$ vector of parameters $\mathbf{\Lambda} = \mathrm{vec}(\mathbf{\Phi}, \mathbf{\Theta})$.

## 1.2 An introduction of goodness-of-fit literature in univariate time series

Goodness of fit in time series models has received great attention in the statistical literature. A number of authors, for example Godfrey (1979), Newbold (1980) and

Poskitt and Tremayne (1980), considered diagnostic checking of univariate linear time series models. The commonly used model is the *ARMA(p,q)* process

$$\phi(B)(X_t - \mu) = \theta(B)\varepsilon_t \ , \tag{1.2}$$

where $\mu = E(X_t)$; and $B$ denotes the backward shift operator $BX_t = X_{t-1}$. Additionally, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ are polynomials of degrees $p$ and $q$, respectively. On the other hand, $\varepsilon_t \sim WN(0, \sigma^2)$. Model (1.2) is a the univariate version of (1.1) (Brockwell and Davis, 1991, Chapter. 3).

Given a finite observed series $(X_1, \ldots, X_n)'$ from model (1.2), the mean $\mu$ can be estimated by the average $\overline{X}_n = \sum_{t=1}^{n} X_t/n$. The parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$ can be estimated by the least squares estimates

$$(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}) = \arg\min_{(\phi,\theta)} \sum_{t>P}^{n} [\varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \overline{X}_n)]^2 \ , \tag{1.3}$$

where the functions $\{\varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \mu) : 1 \leq t \leq n\}$ are defined implicitly by mean of the equations $\phi(B)(X_t - \mu) = \theta(B)\varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$; and the conditions $X_t - \mu \equiv 0 \equiv \varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$, $t \leq 0$. Once that the components of the pair $(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}})$ have been determined, the least squares residuals are defined as $\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}, \overline{X}_n)$. Therefore, for $P < t \leq n$, they can be obtained recursively as

$$\widehat{\varepsilon}_t = (X_t - \overline{X}_n) - \sum_{i=1}^{p} \widehat{\phi}_i (X_{t-i} - \overline{X}_n) - \sum_{j=1}^{q} \widehat{\theta}_j \widehat{\varepsilon}_{t-j} \ . \tag{1.4}$$

The residual autocorrelation function

$$\widehat{r}_k = \frac{\sum_{t>P}^{n-k} \widehat{\varepsilon}_t \widehat{\varepsilon}_{t+k}}{\sum_{t>P}^{n} \widehat{\varepsilon}_t^2} \ , \quad k = 1, \ldots, M \ , \tag{1.5}$$

is the most frequently used statistic in practice. Tests of goodness of fit can be based on the $\widehat{r}_k$. This is because, if the model is adequate, the residuals are uncorrelated approximately. Therefore, if $n >> M$, we may expect that $\widehat{r}_1 \simeq \widehat{r}_2 \simeq \ldots \simeq \widehat{r}_M \simeq 0$. The asymptotic properties of the residual autocorrelations are studied next.

### 1.2.1 The asymptotic distribution of the residual autocorrelation function

The asymptotic distribution of the $\widehat{r}_k$ was first derived by Walker (1952) for autoregressive processes, and for univariate *ARMA* models by Box and Pierce (1970). Let $\mathbf{r} = (r_1, \ldots, r_M)'$ be the $M \times 1$ model counterpart of $\widehat{\mathbf{r}} = (\widehat{r}_1, \ldots, \widehat{r}_M)'$, where

$$r_k = \frac{\sum_{t=1}^{n-k} \varepsilon_t \varepsilon_{t+k}}{\sum_{t=1}^{n} \varepsilon_t^2} \ . \tag{1.6}$$

Since the estimators $(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}, \overline{X}_n)$ are consistent, the standard argument is to approximate the value of $\widehat{r}_k$ by a first order Taylor expansion about the true parameters $(\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$. According to McLeod (1979),

$$\widehat{\mathbf{r}} = \mathbf{r} - \mathbf{X}_M(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) + O_P(\frac{1}{n}) \ , \tag{1.7}$$

where $\boldsymbol{\lambda} = (\boldsymbol{\phi}', \boldsymbol{\theta}')'$ is a $(p+q) \times 1$ vector; $\widehat{\boldsymbol{\lambda}} = (\widehat{\boldsymbol{\phi}}', \widehat{\boldsymbol{\theta}}')'$; and $\mathbf{X}_M$ stands for the $M \times (p+q)$ matrix

$$\mathbf{X}_M = \left(\begin{array}{cccc|cccc}
1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
h_1 & 1 & \cdots & \vdots & l_1 & 1 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\
h_{p-1} & h_{p-2} & \cdots & 1 & \vdots & \vdots & \ddots & \vdots \\
\cline{1-4}
h_p & h_{p-1} & \cdots & h_1 & \vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
\vdots & \vdots & \ddots & \vdots & l_{q-1} & l_{q-2} & \cdots & 1 \\
\cline{5-8}
\vdots & \vdots & \ddots & \vdots & l_q & l_{q-1} & \cdots & l_1 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\
h_{M-1} & h_{M-2} & \cdots & h_{M-p} & l_{M-1} & l_{M-2} & \cdots & l_{M-q}
\end{array}\right) , \tag{1.8}$$

whose entries are given by the coefficients $\{h_r : r \geq 0\}$ of the series $\phi^{-1}(z) = \sum_{r=0}^{\infty} h_r z^r$, where $h_0 = 1$; and the coefficients $\{l_r : r \geq 0\}$ of the series $\theta^{-1}(z) = \sum_{r=0}^{\infty} l_r z^r$, where $l_0 = 1$.

As it is well-known, the asymptotic distribution of the $M \times 1$ random vector $\mathbf{r} = (r_1, \ldots, r_M)'$ is $N_M(\mathbf{0}, n^{-1}\,\mathbf{I}_M)$, where $\mathbf{I}_M$ is the identity matrix of order $M$. On the other hand, $\sqrt{n}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \xrightarrow{D} \mathbf{N}_{p+q}[\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\lambda})]$, where $\mathbf{I}(\boldsymbol{\lambda})$ is the $(p+q) \times (p+q)$ information matrix. Additionally,

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} \\ \mathbf{r} \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \mathbf{I}^{-1}(\boldsymbol{\lambda}) & -\mathbf{I}^{-1}(\boldsymbol{\lambda})\,\mathbf{X}_M' \\ -\mathbf{X}_M\,\mathbf{I}^{-1}(\boldsymbol{\lambda}) & \mathbf{I}_M \end{pmatrix} . \tag{1.9}$$

From (1.7) and (1.9) the large sample distribution of the $M \times 1$ random vector $\widehat{\mathbf{r}}$ is approximately multivariate normal with mean $\mathbf{0}$, and $M \times M$ covariance matrix

$$\mathrm{var}_{\boldsymbol{\lambda}}(\widehat{\mathbf{r}}) = \frac{1}{n}[\mathbf{I}_M - \mathbf{X}_M\,\mathbf{I}^{-1}(\boldsymbol{\lambda})\,\mathbf{X}_M'] \ . \tag{1.10}$$

For more details, see Li (2004, chapter 2).

According to Bruce and Martin (1989), $\mathbf{X}'_M \mathbf{X}_M \to \mathbf{I}(\boldsymbol{\lambda})$ as $M$ goes to infinity. Therefore, for $M$ large enough, we can approximate $\mathbf{I}(\boldsymbol{\lambda}) \simeq \mathbf{X}'_M \mathbf{X}_M$. After replacing $\mathbf{I}^{-1}(\boldsymbol{\lambda})$ by $(\mathbf{X}'_M \mathbf{X}_M)^{-1}$ in (1.10), $n\,\mathrm{var}_{\boldsymbol{\lambda}}(\widehat{\mathbf{r}})$ is close to the orthogonal projection matrix $\mathbf{I}_M - \mathbf{X}_M(\mathbf{X}'_M \mathbf{X}_M)^{-1}\mathbf{X}'_M$. Box and Pierce (1970) exploited this fact, and derived the limit distribution of the portmanteau goodness-of-fit test statistic

$$\widehat{Q}_{BP} = n\,\widehat{\mathbf{r}}'\,\widehat{\mathbf{r}} = n \sum_{k=1}^{M} \widehat{r}_k^2 \ . \tag{1.11}$$

If the model is adequate and $M$ is large enough, $\widehat{Q}_{BP}$ is asymptotically chi-squared distributed with $M-(p+q)$ degrees of freedom.

### 1.2.2 Properties and modifications of the portmanteau test statistic

The portmanteau test is a common diagnostic tool for univariate time series models. However, Davies et al. (1977) argued that $\widehat{Q}_{BP}$ could be in practice too conservative, even for a moderate number of observations $n$. To overcome this problem, Ljung and Box (1978) proposed a modification, called the Ljung-Box statistic, that is given by

$$\widehat{Q}_{LB} = n(n+2) \sum_{k=1}^{M} \frac{\widehat{r}_k^2}{(n-k)} \ . \tag{1.12}$$

The finite sample distribution of statistic $\widehat{Q}_{LB}$ is much closer to that of a chi-squared distribution with $M-(p+q)$ degrees of freedom. Additionally, Li and McLeod (1981) introduced an alternative modification of the form

$$\widehat{Q}_{LM} = \widehat{Q}_{BP} + \frac{M(M+1)}{2n} \ . \tag{1.13}$$

The simulation results in Kheoh and McLeod (1992) found that $\widehat{Q}_{LM}$ and $\widehat{Q}_{LB}$ have almost identical power.

On the other hand, Ljung (1986) obtained that

$$\widehat{Q}_{LB} \overset{D}{\sim} \sum_{i=1}^{M} \rho_i \chi_{1,i}^2 \ , \tag{1.14}$$

where the $\rho_i$ are the eigenvalues of the matrix $n\,\mathrm{var}_{\boldsymbol{\lambda}}(\widehat{\mathbf{r}})$, and the $\chi_{1,i}^2$ are independent $\chi_1^2$ random variables. Battaglia (1990) found an approximate relationship between the power of $\widehat{Q}_{LB}$ and the values of the lag $M$.

### 1.2.3 Other goodness-of-fit tests

Whittle (1952) proposed checking the goodness-of-fit of an *AR(p)* using a likelihood-ratio test for an *AR(p+M)* for sufficiently large *M*. Among nested hypothesis procedures such as likelihood ratio, Wald, and Lagrange multiplier (LM) tests, the LM principle is convenient to derive diagnostic checks. This is because it is consistent and asymptotically optimal against the alternative hypotheses. In the context of testing an *ARMA(p,q)* versus an *ARMA(p+M,q)* model, Newbold (1980) showed that the LM test and the test based on the first $M$ residual autocorrelations are equivalent. Another possible test is based on the Wald-type statistic

$$S = n\,\widehat{\mathbf{r}}'\,\mathrm{var}^{-1}_{\widehat{\boldsymbol{\lambda}}}(\widehat{\mathbf{r}})\,\widehat{\mathbf{r}}\ . \tag{1.15}$$

However, Ljung (1986) showed that the size of this procedure is not entirely correct.

Milhøj (1981) suggested a goodness-of-fit test statistic for time series models that is a frequency domain analogue of the Box-Pierce portmanteau statistic $\widehat{Q}_{BP}$.

**A test based on the residual partial autocorrelation**

Monti (1994) proposed a portmanteau test using the residual partial autocorrelations $\widehat{\pi}_k$, $k = 1, \ldots, M$. Given $\widehat{\boldsymbol{\pi}} = (\widehat{\pi}_1, \ldots, \widehat{\pi}_M)'$, the corresponding statistic is defined as

$$\widehat{Q}_{MT} = n(n+2)\sum_{k=1}^{M}\frac{\widehat{\pi}_k^2}{n-k}\ . \tag{1.16}$$

If the fitted *ARMA* model is adequate, then $\widehat{Q}_{MT}$ is asymptotically $\chi^2_{M-(p+q)}$. Monti (1994) showed by simulation that $\widehat{Q}_{MT}$ is more powerful than $\widehat{Q}_{LB}$, if the order of the moving average is understated. However, Kwan and Wu (1997) found very small differences between the powers of $\widehat{Q}_{MT}$ and $\widehat{Q}_{LB}$.

**A test based on the residual correlation matrix**

Peña and Rodríguez (2002) proposed a test based on the residual correlation matrix of order $M$, defined as

$$\widehat{\mathcal{R}}_M = \begin{bmatrix} 1 & \widehat{r}_1 & \ldots & \widehat{r}_M \\ \widehat{r}_1 & 1 & \ldots & \widehat{r}_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{r}_M & \widehat{r}_{M-1} & \ldots & 1 \end{bmatrix}\ . \tag{1.17}$$

The proposed portmanteau statistic is of form

$$\widehat{D}_M = n[1 - |\widehat{\mathcal{R}}_M|^{1/M}] \ . \tag{1.18}$$

It is well-known that $|\widehat{\mathcal{R}}_M| = |\widehat{\mathcal{R}}_{M-1}|(1-\widehat{R}^2_M)$, where $\widehat{R}^2_M = \widehat{\mathbf{r}}'\widehat{\mathcal{R}}^{-1}_{M-1}\widehat{\mathbf{r}}$ is the square of the multiple correlation coefficient in the least-squares regression of the estimated residuals $\widehat{\varepsilon}_t$ on $\widehat{\varepsilon}_{t-1}, \ldots, \widehat{\varepsilon}_{t-M}$. Iterating this formula for $M$, $M-1$, …, 1, gives the expression $|\widehat{\mathcal{R}}_M| \ = (1-\widehat{R}^2_1)\ldots(1-\widehat{R}^2_M)$.

If the model is correctly identified, $\widehat{D}_M$ is asymptotically distributed as $\sum_{i=1}^M \rho_i \chi^2_{1,i}$, where the $\rho_i$ are now the eigenvalues of $[\mathbf{I}_M - \mathbf{X}_M \mathbf{I}^{-1}(\boldsymbol{\lambda})\mathbf{X}'_M]\mathbf{W}_M$; $\mathbf{W}_M$ is a diagonal matrix with $i-th$ diagonal element $W_i = (M-i+1)/M, \ i = 1, \ldots, M$; and the $\chi^2_{1,i}$ are independent $\chi^2_1$ random variables. The distribution of $\widehat{D}_M$ can be approximated by a gamma distribution $G(\alpha, \beta)$ with parameters $\alpha = (\sum \rho_i)^2/(2\sum \rho_i^2)$ and $\beta = \sum \rho_i^2/(2\sum \rho_i)$. Peña and Rodríguez (2002) showed by simulation that the power of $\widehat{D}_M$ is better than that of either $\widehat{Q}_{LB}$ or $\widehat{Q}_{MT}$. However, posterior empirical studies by Kwan and Wu (2003) and Lin and McLeod (2006) detected some problems with the size and power of the statistic (1.18).

### Other procedures

Portmanteau tests have been long popular as tools for model checking. Kwan and Sim (1996) recommended a procedure based on applying Jenkins' variance stabilizing transformation to the sample autocorrelations. Francq et al. (2005) established that the standard Box-Pierce and Ljung-Box portmanteau tests are not suitable under weak assumptions on the noise. Peña and Rodríguez (2006) proposed a finite sample modification of the test in Peña and Rodríguez (2002). The test statistic is now $\widehat{D}^*_M = -[n/(M+1)]\log|\widehat{\mathcal{R}}_M|$. Depending on the model and sample size, $\widehat{D}^*_M$ is more powerful than the ones by Ljung and Box (1978), Monti (1994), Hong (1996), and Li and McLeod (1981). Lin and McLeod (2008) and Lee and Ng (2010) considered portmanteau tests for time series with infinite variance. In Delgado and Velasco (2010), a class of asymptotically pivotal tests was considered based on quadratic forms of weighted sums of residuals autocorrelations.

## 1.3 An introduction of goodness-of-fit literature in multivariate time series

In contrast to the univariate case, diagnostic checks for multivariate linear time series models are still somewhat less developed. Chitturi (1974) considered the $m \times m$

sample error covariance matrices

$$\mathbf{C}_k = \frac{1}{n}\sum_{t=1}^{n-k} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+k} \ , \quad 0 \leq k \leq n-1 \ , \tag{1.19}$$

where $\{\boldsymbol{\varepsilon}_t : 1 \leq t \leq n\}$ are the errors of model (1.1). The sample version of (1.19) is

$$\widehat{\mathbf{C}}_k = \frac{1}{n}\sum_{t>P}^{n-k} \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}'_{t+k} \ , \quad 0 \leq k \leq n-(P+1) \ , \tag{1.20}$$

where the $\widehat{\boldsymbol{\varepsilon}}_t$, that will be analyzed later in more detail in chapter 2, are the natural $m \times 1$ generalizations of the univariate residuals $\widehat{\varepsilon}_t$ of expression (1.4). Using (1.20), Chitturi (1974) defined also the sequence of $m \times m$ residual autocorrelation matrices

$$\widehat{\mathbf{R}}_k = \widehat{\mathbf{C}}'_k \widehat{\mathbf{C}}_0^{-1} \ , \quad 1 \leq k \leq n-(P+1) \ . \tag{1.21}$$

The statistics in (1.21) can be seen as matrix analogues of the univariate quantities $\widehat{r}_k$ of (1.5). For a *VARMA(p,q)* model, Hosking (1980) derived a portmanteau test statistic of the form

$$\widehat{Q}_H^m = n\sum_{k=1}^{M} \mathrm{tr}(\widehat{\mathbf{C}}'_k \widehat{\mathbf{C}}_0^{-1} \widehat{\mathbf{C}}_k \widehat{\mathbf{C}}_0^{-1}) \ , \tag{1.22}$$

where $M$ satisfies the assumptions in Hosking (1980, p. 607). These lead usually to the choice $M = O(n^{1/2})$. If the model is adequate, $\widehat{Q}_H^m$ is asymptotically chi-squared distributed with $m^2[M-(p+q)]$ degrees of freedom.

Hosking (1980) established that $\widehat{Q}_H^m$ can be written as a function of the entries of the matrices in (1.21). Thus, (1.22) generalizes a goodness-of-fit statistic given earlier by Chitturi (1974) for autoregressive schemes *VAR(p)*. Some equivalent forms of (1.22) are derived by Hosking (1981). Hosking (1980) suggested also using

$$\widehat{Q}_{HM}^m = n^2\sum_{k=1}^{M} \frac{\mathrm{tr}(\widehat{\mathbf{C}}'_k \widehat{\mathbf{C}}_0^{-1} \widehat{\mathbf{C}}_k \widehat{\mathbf{C}}_0^{-1})}{n-k} \ . \tag{1.23}$$

Modification (1.23) is inspired by the structure of the univariate Ljung-Box statistic $\widehat{Q}_{LB}$ of (1.12). Li and McLeod (1981) considered

$$\widehat{Q}_{LM}^m = n\sum_{k=1}^{M} \mathrm{tr}(\widehat{\mathbf{C}}'_k \widehat{\mathbf{C}}_0^{-1} \widehat{\mathbf{C}}_k \widehat{\mathbf{C}}_0^{-1}) + \frac{m^2 M(M+1)}{2n} \ , \tag{1.24}$$

whose behavior seems to improve over that of $\widehat{Q}_H^m$ and $\widehat{Q}_{HM}^m$.

**Other procedures**

Hosking (1981) found that the portmanteau test for *VARMA(p,q)* models can be obtained as a LM test against specific alternative hypotheses. Poskitt and Tremayne (1982) introduced a LM test under a Pitman sequence of alternatives. Chen and Deo (2004) investigated a goodness-of-fit test based on the discrete spectral average estimator. This procedure, that does not require the calculation of residuals from the fitted model, is a frequency domain analogue of Hong (1996). A state-space representation for vector autoregressive moving average models, that enables maximum likelihood estimation, is studied in Paparoditis (2005). Kwan et al. (2005) considered the finite-sample performance of some portmanteau tests for randomness of a time series. Lütkepohl (2006) studied model specification, estimation, model checking, and forecasting for *VARMA(p,q)* processes. Bouhaddioui and Roy (2006) proposed a multivariate extension of Hong (1996). A modified goodness-of-fit portmanteau test is given in Francq and Rassi (2007) in the presence of nonindependent innovations. Finally, Mahdi and McLeod (2012) presented a multivariate extension of the method of Peña and Rodríguez (2002). The null asymptotic distribution of this new statistic is similar in structure to that of $\widehat{D}_M$ in (1.18).

## 1.4 Aims

### 1.4.1 Motivation

Ubierna and Velilla (2007) considered, expanding previous results in Velilla (1994), a goodness-of-fit process for univariate *ARMA(p,q)* models. The basic idea is to start with the process $\{\widehat{W}_n(u) : 0 \leq u \leq 1\}$, where

$$\widehat{W}_n(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-(P+1)} \widehat{r}_k \, \frac{\sin(k\pi u)}{k} \tag{1.25}$$

is a random element in $C[0,1]$, the space of continuous functions in $[0,1]$. According to Durbin (1975, sec. 2), $\{\widehat{W}_n(u) : 0 \leq u \leq 1\}$ converges weakly, as $n \to \infty$, to a centered Gaussian process $\{G(u) : 0 \leq u \leq 1\}$ with covariance function

$$\gamma(u,v) = [\min(u,v) - uv] - \frac{1}{2\pi^2}\,\mathbf{g}(\pi u)'\,\mathbf{I}^{-1}(\boldsymbol{\lambda})\,\mathbf{g}(\pi v)\ , \qquad 0 \leq u, v \leq 1\ , \tag{1.26}$$

where $\mathbf{g}(\pi u) = \int_0^{\pi u}[\,\partial \log f(\omega)/\partial\boldsymbol{\lambda}\,]\,d\omega$ is a vector of size $(p+q)\times 1$ that depends on the normalized spectral density function $f(.)$ of an *ARMA(p,q)* process.

By the continuous mapping theorem, the asymptotic distribution of any continuous functional $H[\widehat{W}_n(u)]$ of the process (1.25) is given by $H[G(u)]$. Since this distribution depends on the unknown parameter vector $\boldsymbol{\lambda} = (\boldsymbol{\phi}', \boldsymbol{\theta}')'$, assessing the significance of an observed value of $H[\widehat{W}_n(u)]$ with $H[G(u)]$ is not feasible. Therefore, Ubierna and Velilla (2007) introduced a modified goodness-of-fit process based on a transformed correlation sequence $\{\widehat{s}_k\}$. After replacing in (1.25) the original $\widehat{r}_k$ by these new $\widehat{s}_k$, it can be seen that convergence is now to the Brownian bridge $\{B(u) : 0 \leq u \leq 1\}$. This has the advantage of using the pivotal distribution of $H[B(u)]$ for assessing significance. For example, considering $H[B(u)] = \sup_{0\leq u\leq 1} |B(u)|$ requires the tables of the standard Kolmogorov-Smirnov statistic. Ubierna and Velilla (2007) showed by simulation that their technique improves over standard methods of goodness-of-fit.

### 1.4.2 Purposes and contributions

The purpose of this thesis is to extend the univariate results in Ubierna and Velilla (2007) to the multivariate case. The basic outline is as follows:

**(a)** In chapter 2, we study the properties of the residual vectors $\widehat{\boldsymbol{\varepsilon}}_t$, and give a linear representation that connects the residual covariance matrices of (1.20) to their model error versions of (1.19). We also derive an explicit form of the information matrix as a limit in the multivariate case, that expands previous univariate results by Bruce and Martin (1989). Some useful asymptotic results are established as well. In addition, we consider an initial goodness-of-fit process $\{W_n^m(u) : 0 \leq u \leq 1\}$ of the form

$$W_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-1} \frac{\mathrm{tr}(\mathbf{R}_k)}{\sqrt{m}}\,\frac{\sin(k\pi u)}{k}\ , \tag{1.27}$$

where $\mathbf{R}_k = \mathbf{C}_k'\mathbf{C}_0^{-1}$. As it turns out, $W_n^m(u) \to_\omega B(u)$ as $n \to \infty$. The process of (1.27) is not feasible. However, it can serve as a building block for goodness-of-fit purposes in *VARMA(p,q)* processes.

**(b)** In chapter 3, the sample version of (1.27),

$$\widehat{W}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k}\ , \tag{1.28}$$

is studied, where $\widehat{\mathbf{R}}_k = \widehat{\mathbf{C}}_k'\widehat{\mathbf{C}}_0^{-1}$ is the matrix of (1.21). We will refer to

$$\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}\ , \quad 1 \leq k \leq n-(P+1)\ , \tag{1.29}$$

as the sequence of adjusted residual traces. When $m = 1$, the statistics of (1.29) coincide with the residual autocorrelations $\widehat{r}_k$ of (1.5). Thus, (1.28) is a generalization of the process by Durbin (1975) in (1.25). The asymptotic behavior of $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ is studied in a similar setting to that considered by Ubierna and Velilla (2007). Under a *VARMA(p,q)* specification, $\widehat{W}_n^m(u)$ converges to a Gaussian process whose parametric covariance function has an structure that resembles that appearing in (1.26).

**(c)** In chapter 4, a modified sequence of adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ is introduced, where $\widehat{\mathbf{S}}_k$ is a properly selected $m \times m$ matrix, $k = p+q+1, \ldots, n-(P+1)$. The construction and properties of the $\widehat{\mathbf{S}}_k$ are analyzed. Replacing in (1.28) the original $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ by the new statistics $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ leads to the modified process

$$\widehat{Z}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=p+q+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}} \frac{\sin(K\pi u)}{K}\ , \tag{1.30}$$

where $K = k - (p+q)$. It is established that $\widehat{Z}_n^m(u)$ converges weakly to the Brownian bridge as $n \to \infty$. Therefore, we suggest assessing goodness-of-fit of a *VARMA(p,q)* model using statistics defined by continuous functionals of the form $H[\widehat{Z}_n^m(u)]$, whose null pivotal distribution is given by $H[B(u)]$. As for instance, the Kolmogorov-Smirnov statistic $H[\widehat{Z}_n^m(u)] = \sup_{0\leq u\leq 1}|\widehat{Z}_n^m(u)|$; and the Cramér-von Mises statistic $H[\widehat{Z}_n^m(u)] = \int_0^1 [\widehat{Z}_n^m(u)]^2 du$.

**(d)** In chapter 5, we investigate the empirical behavior of our technique suggested in chapter 4. Simulation studies are presented, as well as comparisons in both size and power with the statistics $\widehat{Q}_H^m$ in (1.22) by Hosking (1980); and $\widehat{Q}_{LM}^m$ in (1.24) by Li and McLeod (1981). A new version of the cumulative periodogram is defined in terms of the adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ and $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$. Applications in model selection with real data are also given.

Multivariate time series require eventually the use of cumbersome notation. This may obscure the presentation of some specific results. For this reason, technical derivations are often included in the proper additional mathematical appendices.

Some matrices results are frequently used. For reference, they are summarized below. Put $\otimes$ for the Kronecker product of matrices, and $\mathrm{vec}(\mathbf{A})$ for the operator that vectorizes a $p \times q$ matrix $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_q)$ in the form $\mathrm{vec}(\mathbf{A}) = (\mathbf{a}_1', \ldots, \mathbf{a}_q')'$. Then, for matrices $\mathbf{A}$, $\mathbf{X}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ of adequate dimensions:

**(a)** $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$.

**(b)** $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$.

**(c)** $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$.

**(d)** $\mathrm{vec}(\mathbf{AXB}) = (\mathbf{B}' \otimes \mathbf{A})\mathrm{vec}(\mathbf{X})$.

**(e)** $[\mathrm{vec}(\mathbf{A}')]'[\mathrm{vec}(\mathbf{B})] = \mathrm{tr}(\mathbf{AB})$.

These expressions can be found, for example, in Fang and Zhang (1990).

# Chapter 2

# Preliminary results

**Summary.** This chapter presents some preliminary results for a new goodness-of-fit method for *VARMA(p,q)* models. In sections 2.1 and 2.2, we consider maximum likelihood (ML) estimators; and their derivation using an adequate iterative procedure. Section 2.3 contains some essential results relative to the information matrix as a limit. Relations between least squares residuals and true errors are also examined in section 2.4. As an application, the properties of the residual traces are analyzed in section 2.5. An initial error goodness-of-fit process for *VARMA(p,q)* models is introduced. This can be seen as a multivariate extension of those studied by Durlauf (1991) and Anderson (1993). Some useful auxiliary asymptotic results are also presented.

## 2.1 Parameter estimation and residuals

Given $n$ observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from model (1.1), the mean vector $\boldsymbol{\mu}$ can be estimated by the sample mean $\overline{\mathbf{X}}_n = n^{-1} \sum_{t=1}^{n} \mathbf{X}_t$. To estimate the remaining parameters $(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma})$, we consider the collection of $m \times 1$ vectors $\{\boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\mu}) : 1 \leq t \leq n\}$, that are defined recursively using the equations

$$\boldsymbol{\Phi}(B)(\mathbf{X}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(B)\boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\mu}) \ ; \tag{2.1}$$

and the initial conditions $\mathbf{X}_t - \boldsymbol{\mu} \equiv \mathbf{0} \equiv \boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\mu})$, $t \leq 0$. In practice, the ML estimators of $(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma})$ are considered. Following Lütkepohl (2005, sec.12.2), these are obtained by maximizing the Gaussian likelihood function. This problem is equivalent to minimizing the objective function

$$l_n(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma}) = \frac{n}{2} \log(|\boldsymbol{\Sigma}|) + \frac{1}{2} \sum_{t>P}^{n} \boldsymbol{\varepsilon}_t'(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \overline{\mathbf{X}}_n)\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \overline{\mathbf{X}}_n) \ . \tag{2.2}$$

For fixed $\boldsymbol{\Sigma}$, we denote the optimizers of (2.2) as $(\overline{\boldsymbol{\Phi}}(\boldsymbol{\Sigma}), \overline{\boldsymbol{\Theta}}(\boldsymbol{\Sigma}))$. In the univariate case, $\boldsymbol{\Sigma}$ is a scalar parameter $\sigma^2 > 0$. Thus, it is easy to see that $(\overline{\boldsymbol{\Phi}}(\sigma^2), \overline{\boldsymbol{\Theta}}(\sigma^2))$ do not depend on $\sigma^2$. Then, these are the ML estimates of $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$, respectively. However, in the multivariate case $(\overline{\boldsymbol{\Phi}}(\boldsymbol{\Sigma}), \overline{\boldsymbol{\Theta}}(\boldsymbol{\Sigma}))$ depend in general on $\boldsymbol{\Sigma}$. Therefore, the ML estimators $(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\Sigma}})$ of $(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma})$ are of the form $\widehat{\boldsymbol{\Phi}} = \overline{\boldsymbol{\Phi}}(\widehat{\boldsymbol{\Sigma}})$ and $\widehat{\boldsymbol{\Theta}} = \overline{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Sigma}})$, where $\widehat{\boldsymbol{\Sigma}} = \arg\min_{\boldsymbol{\Sigma}} l_n(\overline{\boldsymbol{\Phi}}(\boldsymbol{\Sigma}), \overline{\boldsymbol{\Theta}}(\boldsymbol{\Sigma}), \boldsymbol{\Sigma})$. Optimization of (2.2) must then performed with respect to $(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma})$ simultaneously. This is a complex nonlinear optimization problem, affected by the potentially large number of parameters involved, and that must be solved using an adequate efficient algorithm.

According to Lütkepohl (2005, p. 408), the vector of ML estimators $\widehat{\mathbf{\Lambda}} = \mathrm{vec}(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}})$ is consistent and asymptotically normal for $\mathbf{\Lambda} = \mathrm{vec}(\mathbf{\Phi}, \mathbf{\Theta})$. In particular,

$$\sqrt{n}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}) \xrightarrow{D} N_{m^2(p+q)}[\mathbf{0}, \mathbf{I}^{-1}(\mathbf{\Lambda})] \ , \tag{2.3}$$

where $\mathbf{I}(\mathbf{\Lambda})$ is the $m^2(p+q) \times m^2(p+q)$ information matrix.

Once that $\widehat{\mathbf{\Lambda}} = \mathrm{vec}(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}})$ has been determined, the $m \times 1$ residual vectors $\widehat{\boldsymbol{\varepsilon}}_t$, $t = 1, \ldots, n$, are defined recursively using (2.1):

$$\widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{\varepsilon}_t(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}}, \overline{\mathbf{X}}_n) = (\mathbf{X}_t - \overline{\mathbf{X}}_n) - \sum_{i=1}^{p} \widehat{\mathbf{\Phi}}_i(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n) - \sum_{j=1}^{q} \widehat{\mathbf{\Theta}}_j \widehat{\boldsymbol{\varepsilon}}_{t-j} \ , \quad t = 1, \ldots, n \ , \tag{2.4}$$

with the usual conditions $\mathbf{X}_t - \overline{\mathbf{X}}_n \equiv \mathbf{0} \equiv \widehat{\boldsymbol{\varepsilon}}_t$, for $t \leq 0$. In practice, only residual vectors for $t > P = \max(p, q)$ are considered.

## 2.2 Computation of the ML estimates

### 2.2.1 *VAR(p)* case

In a *VAR(p)* model, $q = 0$. Then, $P = \max(p, q) = p$. The functions $\boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \boldsymbol{\mu})$ coincide, for $t > P$, with the true error vectors $\boldsymbol{\varepsilon}_t$:

$$\boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \boldsymbol{\mu}) = (\mathbf{X}_t - \boldsymbol{\mu}) - \sum_{i=1}^{P} \mathbf{\Phi}_i(\mathbf{X}_{t-i} - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t \ . \tag{2.5}$$

The objective function (2.2) is of the form

$$l_n(\mathbf{\Phi}, \mathbf{\Sigma}) = \frac{n}{2} \log(|\mathbf{\Sigma}|) + \frac{1}{2} \sum_{t>P}^{n} \boldsymbol{\varepsilon}'_t(\mathbf{\Phi}, \overline{\mathbf{X}}_n) \mathbf{\Sigma}^{-1} \boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \overline{\mathbf{X}}_n) \ . \tag{2.6}$$

The equations $\partial l_n(\mathbf{\Phi}, \mathbf{\Sigma}) / \partial \mathbf{\Sigma} = \mathbf{0}$ lead to the solution

$$\mathbf{\Sigma}_n(\mathbf{\Phi}) = \frac{1}{n} \sum_{t>P}^{n} \boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \overline{\mathbf{X}}_n) \boldsymbol{\varepsilon}'_t(\mathbf{\Phi}, \overline{\mathbf{X}}_n) \ . \tag{2.7}$$

Therefore, the computational ML problem reduces to estimating the parameters $\mathbf{\Phi}$. The conditions $\partial l_n(\mathbf{\Phi}, \mathbf{\Sigma}) / \partial \mathrm{vec}(\mathbf{\Phi}_i)' = \mathbf{0}$, $i = 1, \ldots, p$, are equivalent to

$$\sum_{t>P}^{n} \boldsymbol{\varepsilon}'_t(\mathbf{\Phi}, \overline{\mathbf{X}}_n) \mathbf{\Sigma}^{-1} \mathbf{D}_{t,i} = \mathbf{0} \ , \quad i = 1, \ldots, p \ , \tag{2.8}$$

where

$$\mathbf{D}_{t,i} = \frac{\partial \boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \overline{\mathbf{X}}_n)}{\partial \mathrm{vec}(\mathbf{\Phi}_i)'} = -[(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n)' \otimes \mathbf{I}_m] \ , \quad i = 1, \ldots, p \ .$$

After some algebra, (2.8) leads to the exact orthogonality conditions

$$\sum_{t>P}^{n} \boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \overline{\mathbf{X}}_n)(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n)' = \mathbf{0} \ , \quad i = 1, \ldots, p \ , \tag{2.9}$$

that are independent of $\boldsymbol{\Sigma}$.

Using expression (2.5) in (2.9), the estimates $\widehat{\boldsymbol{\Phi}}_i$, $i = 1, \ldots, p$, satisfy the normal or Yule-Walker equations

$$\sum_{t>P}^{n}(\mathbf{X}_t - \overline{\mathbf{X}}_n)(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n)' = \sum_{j=1}^{p} \boldsymbol{\Phi}_j [\sum_{t>P}^{n}(\mathbf{X}_{t-j} - \overline{\mathbf{X}}_n)(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n)'] \ , \quad i = 1, \ldots, p \ . \tag{2.10}$$

In particular, for a *VAR(1)* process the Yule-Walker estimate of $\boldsymbol{\Phi}_1$ is

$$\widehat{\boldsymbol{\Phi}}_1 = [\sum_{t>p}^{n}(\mathbf{X}_t - \overline{\mathbf{X}}_n)(\mathbf{X}_{t-1} - \overline{\mathbf{X}}_n)'][\sum_{t>p}^{n}(\mathbf{X}_{t-1} - \overline{\mathbf{X}}_n)(\mathbf{X}_{t-1} - \overline{\mathbf{X}}_n)']^{-1} \ . \tag{2.11}$$

### 2.2.2 *VARMA(p,q)* case

In the *VARMA(p,q)* case, the orthogonality conditions analogue to (2.8) depend on $\boldsymbol{\Sigma}$, and cannot be solved explicitly. Therefore, the use of an adequate algorithm is required. See for example the proposals in Hillmer and Tiao (1979), Nicholls and Hall (1979), Mauricio (1995), and Reinsel (1997, chapter 5).

Lütkepohl (2005, sec.12.3) provided a possible iterative numerical strategy for computing the ML estimates $(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\Sigma}})$. Given parameters $\boldsymbol{\Lambda} = \mathrm{vec}(\boldsymbol{\Phi}, \boldsymbol{\Theta})$, consider the $m \times m^2$ matrices of partial derivatives $\mathbf{D}_{t,i} = \partial \boldsymbol{\varepsilon}_t(\boldsymbol{\Lambda}, \overline{\mathbf{X}}_n)/\partial \mathrm{vec}(\boldsymbol{\Phi}_i)'$, $i = 1, \ldots, p$; and $\mathbf{E}_{t,j} = \partial \boldsymbol{\varepsilon}_t(\boldsymbol{\Lambda}, \overline{\mathbf{X}}_n)/\partial \mathrm{vec}(\boldsymbol{\Theta}_j)'$, $j = 1, \ldots, q$. These can be determined as follows

$$\mathbf{D}_{t,i} = -\sum_{r=0}^{t-i-1} \mathbf{L}_r [(\mathbf{X}_{t-i-r} - \overline{\mathbf{X}}_n)' \otimes \mathbf{I}_m] \ , \quad i = 1, \ldots, p \ ; \tag{2.12}$$

$$\mathbf{E}_{t,j} = -\sum_{r=0}^{t-j-1} \mathbf{L}_r [\boldsymbol{\varepsilon}'_{t-j-r}(\boldsymbol{\Lambda}, \overline{\mathbf{X}}_n) \otimes \mathbf{I}_m] \ , \quad j = 1, \ldots, q \ , \tag{2.13}$$

where the $m \times m$ matrices $\{\mathbf{L}_r : r \geq 0\}$ are the coefficients of the series expansion

$$\boldsymbol{\Theta}^{-1}(z) = \sum_{r=0}^{\infty} \mathbf{L}_r z^r \ . \tag{2.14}$$

It can be also checked that

$$\frac{\partial^2 l_n(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)}{\partial \mathbf{\Lambda} \partial \mathbf{\Lambda}'} = \sum_{t>P}^{n} [\boldsymbol{\varepsilon}_t'(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)\mathbf{\Sigma}^{-1} \otimes \mathbf{I}_{m^2(p+q)}] \partial \mathrm{vec}[\partial \boldsymbol{\varepsilon}_t'(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)/\partial \mathbf{\Lambda}] \partial \mathbf{\Lambda}' + \\ + \sum_{t>P}^{n} [\boldsymbol{\varepsilon}_t'(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)/\partial \mathbf{\Lambda}] \mathbf{\Sigma}^{-1} [\partial \boldsymbol{\varepsilon}_t(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)/\partial \mathbf{\Lambda}'] \ . \tag{2.15}$$

From Lütkepohl (2005, sec.12.3.5), the expected value of the first summand of (2.15) is approximately $\mathbf{0}$. Accordingly, the matrix $\mathbf{I}_n(\mathbf{\Lambda}) = E[\partial^2 l_n(\mathbf{\Phi}, \mathbf{\Theta}, \mathbf{\Sigma})/\partial \mathbf{\Lambda} \partial \mathbf{\Lambda}']$ can be approximated by the second summand in (2.15). On the other hand,

$$\frac{\partial l_n(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)}{\partial \mathbf{\Lambda}} = \sum_{t>P}^{n} \frac{\partial \boldsymbol{\varepsilon}_t'(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)}{\partial \mathbf{\Lambda}} \mathbf{\Sigma}^{-1} \boldsymbol{\varepsilon}_t(\mathbf{\Lambda}, \overline{\mathbf{X}}_n) \ . \tag{2.16}$$

Notice that the structure of (2.16) depends on the matrices $\mathbf{D}_{t,i} = \partial \boldsymbol{\varepsilon}_t(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)/\partial \mathrm{vec}(\mathbf{\Phi}_i)'$, $i = 1, \ldots, p$; $\mathbf{E}_{t,j} = \partial \boldsymbol{\varepsilon}_t(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)/\partial \mathrm{vec}(\mathbf{\Theta}_j)'$, $j = 1, \ldots, q$; and $\mathbf{\Sigma}$.

With all these elements, the iterative procedure is as follows:

Step 1. Determine residuals $\boldsymbol{\varepsilon}_t^0 = \boldsymbol{\varepsilon}_t(\mathbf{\Lambda}^0, \overline{\mathbf{X}}_n)$, where $\mathbf{\Lambda}^0 = \mathrm{vec}(\mathbf{\Phi}^0, \mathbf{\Theta}^0)$ is a set of preliminary estimates. Lütkepohl (2005, sec.12.3.4) gives some recommendations for choosing $\mathbf{\Lambda}^0$. As in (2.7), the starting value for $\mathbf{\Sigma}$ is the $m \times m$ matrix

$$\mathbf{\Sigma}^0 = \frac{1}{n} \sum_{t>P}^{n} \boldsymbol{\varepsilon}_t^0 \boldsymbol{\varepsilon}_t^{0\prime} \ .$$

Step 2. For $k = 0, 1, \ldots,$ perform the scoring type iteration

$$\mathbf{\Lambda}^{k+1} = \mathbf{\Lambda}^k - s^k \mathbf{I}_n^{k^{-1}} [\frac{\partial l_n(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)}{\partial \mathbf{\Lambda}} |_{\mathbf{\Lambda} = \mathbf{\Lambda}^k}] \ , \tag{2.17}$$

where $\mathbf{\Lambda}^k = \mathrm{vec}(\mathbf{\Phi}^k, \mathbf{\Theta}^k)$ is the estimate in the $k$th stage, and $s^k$ is a step length that may be or not equal to 1. The use of a proper scaling factor $s^k$ is advisable to avoid surpassing the point of the minimum (Reinsel, 1997, p.127). On the other hand, $\mathbf{I}_n^k$ is an approximation to the matrix $\mathbf{I}_n(\mathbf{\Lambda}^k)$ of the form

$$\mathbf{I}_n^k = \sum_{t>P}^{n} \frac{\partial \boldsymbol{\varepsilon}_t'(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)}{\partial \mathbf{\Lambda}} |_{\mathbf{\Lambda} = \mathbf{\Lambda}^k} \mathbf{\Sigma}^{k^{-1}} \frac{\partial \boldsymbol{\varepsilon}_t(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)}{\partial \mathbf{\Lambda}'} |_{\mathbf{\Lambda} = \mathbf{\Lambda}^k} \ ,$$

where $\mathbf{\Sigma}^k = \sum_{t>P}^{n} \boldsymbol{\varepsilon}_t^k \boldsymbol{\varepsilon}_t^{k\prime}/n$, and $\boldsymbol{\varepsilon}_t^k = \boldsymbol{\varepsilon}_t(\mathbf{\Lambda}^k, \overline{\mathbf{X}}_n)$. Finally,

$$\frac{\partial l_n(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)}{\partial \mathbf{\Lambda}} |_{\mathbf{\Lambda} = \mathbf{\Lambda}^k} = \sum_{t>P}^{n} \frac{\partial \boldsymbol{\varepsilon}_t'(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)}{\partial \mathbf{\Lambda}} |_{\mathbf{\Lambda} = \mathbf{\Lambda}^k} \mathbf{\Sigma}^{k^{-1}} \boldsymbol{\varepsilon}_t^k \ .$$

After performing the $k$th iteration using (2.17), it follows from well-known results relative to the behavior of the Gaussian likelihood that, after multiplying by $2/n$, the new minimum of the objective function (2.2) is equal to $\log(|\mathbf{\Sigma}^{k+1}|) + m$, where $\mathbf{\Sigma}^{k+1} = \sum_{t>P}^{n} \boldsymbol{\varepsilon}_t^{k+1} \boldsymbol{\varepsilon}_t^{(k+1)'}/n$.

The above method is iterated until convergence is reached, according to the usual numerical criteria. The maximum likelihood estimates $(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}}, \widehat{\mathbf{\Sigma}})$ are such that

$$\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{C}}_0 = \frac{1}{n} \sum_{t>P}^{n} \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t' \ , \tag{2.18}$$

where the $\widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{\varepsilon}_t(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}}, \overline{\mathbf{X}}_n)$ are the residuals of (2.4).

## 2.3 The information matrix as a limit

From (2.3), the asymptotic behavior of the ML estimators $\widehat{\mathbf{\Lambda}} = \mathrm{vec}(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}})$ depends on the information matrix $\mathbf{I}(\mathbf{\Lambda})$ . The latter array can be obtained as a limit. This result will be important later for the diagnostic checks presented in this thesis.

Consider the $m \times m$ coefficients of the series expansions $\mathbf{\Phi}^{-1}(z)\mathbf{\Theta}(z) = \sum_{j=0}^{\infty} \mathbf{\Omega}_j z^j$ and $\mathbf{\Theta}^{-1}(z) = \sum_{j=0}^{\infty} \mathbf{L}_j z^j$, where $\mathbf{\Omega}_0 = \mathbf{L}_0 = \mathbf{I}_m$. Define also the collection of matrices $\mathbf{G}_k = \sum_{j=0}^{k} (\mathbf{\Sigma}\mathbf{\Omega}_j' \otimes \mathbf{L}_{k-j})$, and $\mathbf{F}_k = \mathbf{\Sigma} \otimes \mathbf{L}_k$, $k \geq 0$. Put $\mathbf{G}_k = \mathbf{F}_k = \mathbf{0}$ for $k < 0$. Construct also the sequence of $km^2 \times m^2(p+q)$ matrices $\mathbf{Z}_k = (\mathbf{X}_k, \mathbf{Y}_k)$, $k \geq 1$, where

$$\mathbf{X}_k = \begin{pmatrix} \mathbf{G}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{G}_1 & \mathbf{G}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{G}_2 & \mathbf{G}_1 & \mathbf{G}_0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{k-1} & \mathbf{G}_{k-2} & \mathbf{G}_{k-3} & \cdots & \mathbf{G}_{k-p} \end{pmatrix} , \tag{2.19}$$

and

$$\mathbf{Y}_k = \begin{pmatrix} \mathbf{F}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}_1 & \mathbf{F}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}_2 & \mathbf{F}_1 & \mathbf{F}_0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}_{k-1} & \mathbf{F}_{k-2} & \mathbf{F}_{k-3} & \cdots & \mathbf{F}_{k-q} \end{pmatrix} . \tag{2.20}$$

Write $\mathbf{\Lambda} = \mathrm{vec}(\mathbf{\Phi}, \mathbf{\Theta}) = [\lambda_1, \ldots, \lambda_{m^2(p+q)}]'$. Define also the function $\mathbf{A}(\omega, \lambda_i) = \mathbf{k}^{-1}(\omega, \mathbf{\Lambda})\,[\partial \mathbf{k}(\omega, \mathbf{\Lambda})/\partial \lambda_i]$, $i = 1, \ldots, m^2(p+q)$, where $\mathbf{k}(\omega, \mathbf{\Lambda}) = \mathbf{\Phi}^{-1}(e^{i\omega})\mathbf{\Theta}(e^{i\omega})$.

Following Dunsmuir and Hannan (1976, sec. 4), the $(i,j)$-th entry of the information matrix $\mathbf{I}(\mathbf{\Lambda}) = [I_{i,j}(\mathbf{\Lambda})\colon\ i,j = 1,\ldots,m^2(p+q)]$ is of the form

$$I_{i,j}(\mathbf{\Lambda}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\left[\mathbf{A}(\omega,\lambda_i)\mathbf{\Sigma}\mathbf{A}^*(\omega,\lambda_j)\mathbf{\Sigma}^{-1}\right] d\omega \ , \tag{2.21}$$

where $^*$ denotes the conjugate transpose. Expression (2.21) generalizes that of the univariate case. See e.g. Brockwell and Davis (1991, chapter 8). However, the noise covariance matrix $\mathbf{\Sigma}$ does not cancel in (2.21). Therefore, unlike the univariate case, the information matrix for $m > 1$ is not scale free.

It is convenient to partition

$$\mathbf{I}(\mathbf{\Lambda}) = \begin{pmatrix} \mathbf{I}_{11}(\mathbf{\Lambda}) & \mathbf{I}_{12}(\mathbf{\Lambda}) \\ \mathbf{I}_{21}(\mathbf{\Lambda}) & \mathbf{I}_{22}(\mathbf{\Lambda}) \end{pmatrix} \ , \tag{2.22}$$

where the notation indicates that: $\mathbf{I}_{11}(\mathbf{\Lambda})$ is a $m^2p \times m^2p$ matrix that collects together all the terms of the form (2.21) corresponding to $\lambda_i$ and $\lambda_j$ in the set of autoregressive parameters of model (1.1), $\mathbf{\Phi}_r = (\phi_{jk,r} : j,k = 1,\ldots,m)$, $r = 1, \ldots, p$; $\mathbf{I}_{12}(\mathbf{\Lambda})$ is a $m^2p \times m^2q$ matrix that contains the terms (2.21) when $\lambda_i$ is autoregressive and $\lambda_j$ belongs to the family of moving average parameters of model (1.1), $\mathbf{\Theta}_s = (\theta_{jk,s} : j,k = 1,\ldots,m)$, $s = 1, \ldots, q$; and $\mathbf{I}_{21}(\mathbf{\Lambda})$ and $\mathbf{I}_{22}(\mathbf{\Lambda})$ are defined similarly.

After some algebra (see Appendix 2.1), it can be obtained that, for $r, R = 1,\ldots,p$; and $s, S = 1,\ldots,q$:

**(a)** The $(r,R)$ block of $\mathbf{I}_{11}(\mathbf{\Lambda})$ is $\sum_{k=0}^{\infty} \mathbf{G}'_{k-r}(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{G}_{k-R}$.

**(b)** The $(r,s)$ block of $\mathbf{I}_{12}(\mathbf{\Lambda})$ is $\sum_{k=0}^{\infty} \mathbf{G}'_{k-r}(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{F}_{k-s}$.

**(c)** The $(s,S)$ block of $\mathbf{I}_{22}(\mathbf{\Lambda})$ is $\sum_{k=0}^{\infty} \mathbf{F}'_{k-s}(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{F}_{k-S}$.

Put now

$$\mathbf{\Xi}'_k = (\mathbf{\Sigma}^{-1/2}\otimes\mathbf{\Sigma}^{-1/2})(\mathbf{G}_{k-1},\ldots,\mathbf{G}_{k-p}\,;\mathbf{F}_{k-1},\ldots,\mathbf{F}_{k-q}) \tag{2.23}$$

for the $k$th $m^2 \times m^2(p+q)$ row block of the matrix $\mathcal{W}^{-1/2}\mathbf{Z}_M$, $k = 1, \ldots, M$, where $\mathcal{W} = \mathbf{I}_M \otimes \mathbf{\Sigma} \otimes \mathbf{\Sigma}$. Taking into account the expressions in **(a)**, **(b)**, and **(c)** above, the information matrix can be obtained finally as a limit in the form

$$\mathbf{Z}'_M\mathcal{W}^{-1}\mathbf{Z}_M = \sum_{k=1}^{M} \mathbf{\Xi}_k\mathbf{\Xi}'_k \longrightarrow \mathbf{I}(\mathbf{\Lambda}) \ , \quad \text{as } M \to \infty \ . \tag{2.24}$$

Similar results are obtained by Klein and Spreij (2004). The asymptotic representation (2.24) generalizes the univariate findings of Bruce and Martin (1989). As it will be studied later in chapter 5, the matrices $\mathbf{G}_k$ and $\mathbf{F}_k$ that appear in expression (2.23) are exponentially bounded. Thus, only the first few terms of the sum in (2.24) are really needed to obtain an adequate approximation of the information matrix $\mathbf{I}(\mathbf{\Lambda})$.

## 2.4 Linear relations

From Lemma 1 in Hosking (1980, p. 603), a Taylor's expansion of $\widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{\varepsilon}_t(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Theta}}, \overline{\mathbf{X}}_n)$ about the true parameter values $(\boldsymbol{\Phi}, \boldsymbol{\Theta})$ leads to the following linear relation between the residuals and the errors of model (1.1):

$$\widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{\varepsilon}_t - \sum_{i=1}^{p}\sum_{r=0}^{\infty}\mathbf{L}_r(\widehat{\boldsymbol{\Phi}}_i - \boldsymbol{\Phi}_i)(\mathbf{X}_{t-i-r} - \boldsymbol{\mu}) - \sum_{j=1}^{q}\sum_{r=0}^{\infty}\mathbf{L}_r(\widehat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j)\boldsymbol{\varepsilon}_{t-j-r} + O_P(\frac{1}{n}) \ . \tag{2.25}$$

On the other hand, from Lemma 2 in Hosking (1980, p. 603) the relation between the residual and error covariance matrices of (1.20) and (1.19) is

$$\widehat{\mathbf{C}}'_k = \mathbf{C}'_k - \sum_{i=1}^{p}\sum_{r=0}^{k-i}\mathbf{L}_{k-i-r}(\widehat{\boldsymbol{\Phi}}_i - \boldsymbol{\Phi}_i)\boldsymbol{\Omega}_r\boldsymbol{\Sigma} - \sum_{j=1}^{q}\mathbf{L}_{k-j}(\widehat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j)\boldsymbol{\Sigma} + O_P(\frac{1}{n}) \ . \tag{2.26}$$

From the identity of (2.18), when $k = 0$ relation (2.26) leads to $\widehat{\boldsymbol{\Sigma}} = \mathbf{C}_0 + O_P(1/n)$. By the law of the large numbers, $\mathbf{C}_0 \xrightarrow{P} \boldsymbol{\Sigma}$. Therefore, $\widehat{\boldsymbol{\Sigma}}$ is a consistent estimator for the covariance matrix $\boldsymbol{\Sigma}$ of the errors of model (1.1).

### 2.4.1 Consequences

After taking vecs in both sides of (2.26) it follows that, for each $M \geq 1$,

$$\begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{C}}'_1) \\ \mathrm{vec}(\widehat{\mathbf{C}}'_2) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{C}}'_M) \end{pmatrix} = \begin{pmatrix} \mathrm{vec}(\mathbf{C}'_1) \\ \mathrm{vec}(\mathbf{C}'_2) \\ \vdots \\ \mathrm{vec}(\mathbf{C}'_M) \end{pmatrix} - \mathbf{Z}_M \mathrm{vec}[(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Theta}}) - (\boldsymbol{\Phi}, \boldsymbol{\Theta})] + O_P(\frac{1}{n}) \ , \tag{2.27}$$

where the $m^2M \times m^2(p+q)$ matrix $\mathbf{Z}_M = (\mathbf{X}_M, \mathbf{Y}_M)$ is as defined by expressions (2.19) and (2.20) in section 2.3 for $\mathbf{X}_M$ and $\mathbf{Y}_M$, respectively. On the other hand, according to Hosking (1980), the following approximate orthogonality condition holds for $M = O(\sqrt{n})$

$$\mathbf{Z}'_M \mathcal{W}^{-1} \begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{C}}'_1) \\ \mathrm{vec}(\widehat{\mathbf{C}}'_2) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{C}}'_M) \end{pmatrix} = O_P(\frac{1}{n}) \ . \tag{2.28}$$

Define now the $Mm^2 \times Mm^2$ block diagonal matrix $\mathbf{W} = \mathrm{diag}(\mathbf{C}_0 \otimes \mathbf{C}_0, \overset{(M)}{\cdots}, \mathbf{C}_0 \otimes \mathbf{C}_0) = \mathbf{I}_M \otimes \mathbf{C}_0 \otimes \mathbf{C}_0$. The residual counterpart is $\widehat{\mathbf{W}} = \mathbf{I}_M \otimes \widehat{\boldsymbol{\Sigma}} \otimes \widehat{\boldsymbol{\Sigma}}$. Using (2.27) and (2.28), it can be written after some algebra (Hosking, 1980)

$$\widehat{\mathbf{W}}^{-1/2}\begin{pmatrix}\mathrm{vec}(\widehat{\mathbf{C}}_1')\\ \mathrm{vec}(\widehat{\mathbf{C}}_2')\\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{C}}_M')\end{pmatrix} = (\mathbf{I}_{Mm^2} - \mathbf{P}_M)\mathbf{W}^{-1/2}\begin{pmatrix}\mathrm{vec}(\mathbf{C}_1')\\ \mathrm{vec}(\mathbf{C}_2')\\ \vdots \\ \mathrm{vec}(\mathbf{C}_M')\end{pmatrix} + O_P(\frac{1}{n}) \ , \tag{2.29}$$

where $\mathbf{P}_M = \mathcal{W}^{-1/2}\mathbf{Z}_M(\mathbf{Z}_M'\mathcal{W}^{-1}\mathbf{Z}_M)^{-1}\mathbf{Z}_M'\mathcal{W}^{-1/2}$ is the $Mm^2 \times Mm^2$ orthogonal projection matrix onto the subspace spanned by the columns of $\mathcal{W}^{-1/2}\mathbf{Z}_M$.

The random behavior of the left-hand side of (2.29) is essentially described by the projection of a random vector that, as established below in section 2.6.2, is asymptotically $N_{Mm^2}(\mathbf{0}, \mathbf{I}_{Mm^2})$. In other words,

$$\sqrt{n}\widehat{\mathbf{W}}^{-1/2}\begin{pmatrix}\mathrm{vec}(\widehat{\mathbf{C}}_1')\\ \mathrm{vec}(\widehat{\mathbf{C}}_2')\\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{C}}_M')\end{pmatrix} \overset{D}{\cong} N_{Mm^2}(\mathbf{0}, \mathbf{I}_{Mm^2} - \mathbf{P}_M) \ . \tag{2.30}$$

### 2.4.2 Relation with the univariate case

Relation (2.27) is similar to that given by McLeod (1979) in (1.7) for $m = 1$. Box and Pierce (1970, section 5) consider a univariate version of (2.29) given by

$$\begin{pmatrix}\widehat{r}_1\\ \widehat{r}_2\\ \vdots \\ \widehat{r}_M\end{pmatrix} = [\mathbf{I}_M - \mathbf{A}_M(\mathbf{A}_M'\mathbf{A}_M)^{-1}\mathbf{A}_M']\begin{pmatrix}r_1\\ r_2\\ \vdots \\ r_M\end{pmatrix} + O_P(\frac{1}{n}) \ , \tag{2.31}$$

where

$$\mathbf{A}_M = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_1 & 1 & \cdots & 0 \\ a_2 & a_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{M-1} & a_{M-2} & \cdots & a_{M-(p+q)} \end{pmatrix} \tag{2.32}$$

is a $M \times (p+q)$ matrix that depends on the coefficients $\{a_r : r \geq 0\}$ of the series expansion $a(z) = [\phi(z)\theta(z)]^{-1} = [\theta(z)\phi(z)]^{-1} = \sum_{r=0}^{\infty} a_r z^r$.

Relations (2.29) and (2.31) are similar in spirit. However, there are some important differences. The matrix $\mathbf{A}_M$ of (2.32) does not depend on $\sigma^2$, and it has a ladder structure. On the contrary, as indicated by the expressions in section 2.3 for $\mathbf{X}_M$

and $\mathbf{Y}_M$, the array $\mathcal{W}^{-1/2}\mathbf{Z}_M$ and thus the projection matrix $\mathbf{P}_M$ depend on $\mathbf{\Sigma}$. In addition, $\mathcal{W}^{-1/2}\mathbf{Z}_M$ has an individual ladder structure in each of the two parts of the $m^2 \times m^2(p+q)$ row block matrices $\mathbf{\Xi}'_k$ of (2.23).

In the univariate case, the matrix $\mathcal{W}^{-1/2}\mathbf{Z}_M$ reduces to the matrix $\mathbf{X}_M$ considered by McLeod (1979) in expression (1.8). This is because, for $m = 1$, it is easy to check that $\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2} = 1/\sigma^2$; $\mathbf{G}_k = \sigma^2 h_k$; and $\mathbf{F}_k = \sigma^2 l_k$, $k \geq 0$. The equivalence between the subspaces spanned by this $\mathbf{X}_M$ in (1.8) and the matrix $\mathbf{A}_M$ in (2.32) is a consequence of the commutativity identity $\phi(B)\theta(B) = \theta(B)\phi(B)$. For conciseness, details are omitted. However, following Hosking (1980, section 8), this argument is not valid in the multivariate context, since matrix multiplication is not in general commutative. See also the comments on this issue by Pierce (1970).

## 2.5 Properties of the adjusted residual traces

The linear relation (2.29) is highly dimensional. This is because its left-hand side is a $Mm^2 \times 1$ vector. Therefore, unless a very large sample is available, (2.29) is not entirely convenient in applications. A natural question then is how to simplify this representation so that, as in expression (2.31) by Box and Pierce (1970), both sides of (2.29) become of the order $M \times 1$. A possible solution is discussed next.

Define the $m^2 \times 1$ vector

$$\mathbf{a}_m = \text{vec}(\mathbf{I}_m)/\sqrt{m} \ . \tag{2.33}$$

From the matrix properties listed at the end of section 1.4.2, it follows that $\mathbf{a}'_m\mathbf{a}_m = \text{tr}(\mathbf{I}_m)/m = 1$. Thus $\mathbf{a}_m$ in (2.33) is a unit vector. Next result gives several representations for the adjusted residual traces of expression (1.29) of section 1.4.2,

$$\text{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m} \ , \quad 1 \leq k \leq n-(P+1) \ ,$$

where $\widehat{\mathbf{R}}_k = \widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1}$ is the matrix (1.21) of Chitturi (1974).

**Lemma 2.5.1** The adjusted residual traces of (1.29) can be written

$$\text{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m} = \mathbf{a}'_m\text{vec}(\widehat{\mathbf{R}}_k) \ , \quad 1 \leq k \leq n-(P+1) \ . \tag{2.34}$$

On the other hand,

$$\text{tr}(\widehat{\mathbf{R}}_k) = \text{tr}(\widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1}) = \text{tr}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1/2}) \ . \tag{2.35}$$

Hence, for $1 \leq k \leq n-(P+1)$,

$$\text{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m} = \mathbf{a}'_m\text{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1/2}) = \mathbf{a}'_m(\widehat{\mathbf{\Sigma}}^{-1/2} \otimes \widehat{\mathbf{\Sigma}}^{-1/2})\text{vec}(\widehat{\mathbf{C}}'_k) \ . \tag{2.36}$$

**Proof.** The proof follows form standard properties of the trace operator, and the matrix properties listed at the end of section 1.4.2. ■

From (2.36) it is obtained that

$$\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{R}}_1) \\ \mathrm{tr}(\widehat{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{R}}_M) \end{pmatrix} = (\mathbf{I}_M \otimes \mathbf{a}'_m)\widehat{\mathbf{W}}^{-1/2}\begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{C}}'_1) \\ \mathrm{vec}(\widehat{\mathbf{C}}'_2) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{C}}'_M) \end{pmatrix} , \tag{2.37}$$

where $\widehat{\mathbf{W}} = \mathbf{I}_M \otimes \widehat{\boldsymbol{\Sigma}} \otimes \widehat{\boldsymbol{\Sigma}}$. Hence, from (2.29) it follows that

$$\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{R}}_1) \\ \mathrm{tr}(\widehat{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{R}}_M) \end{pmatrix} = (\mathbf{I}_M \otimes \mathbf{a}'_m)(\mathbf{I}_{Mm^2} - \mathbf{P}_M)\mathbf{W}^{-1/2}\begin{pmatrix} \mathrm{vec}(\mathbf{C}'_1) \\ \mathrm{vec}(\mathbf{C}'_2) \\ \vdots \\ \mathrm{vec}(\mathbf{C}'_M) \end{pmatrix} + O_P(\frac{1}{n}) \ . \tag{2.38}$$

Proceeding as in expression (2.30) and taking into account the identity $(\mathbf{I}_M \otimes \mathbf{a}'_m)(\mathbf{I}_M \otimes \mathbf{a}_m) = \mathbf{I}_M \otimes \mathbf{a}'_m\mathbf{a}_m = \mathbf{I}_M \otimes 1 = \mathbf{I}_M$, the large sample distribution of the left-hand side of (2.38) can be characterized as

$$\sqrt{n}\,[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{R}}_1) \\ \mathrm{tr}(\widehat{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{R}}_M) \end{pmatrix}] \cong N_M[\mathbf{0}, \mathbf{I}_M - (\mathbf{I}_M \otimes \mathbf{a}'_m)\mathbf{P}_M(\mathbf{I}_M \otimes \mathbf{a}_m)] \ . \tag{2.39}$$

The diagonal entries of the covariance matrix in (2.39) are of the form

$$1 - \mathbf{a}'_m\boldsymbol{\Xi}'_k(\mathbf{Z}'_M\mathcal{W}^{-1}\mathbf{Z}_M)^{-1}\boldsymbol{\Xi}_k\mathbf{a}_m \ , \tag{2.40}$$

where, as considered in (2.23), $\boldsymbol{\Xi}'_k$ is the $k$th $m^2 \times m^2(p+q)$ row block of the matrix $\mathcal{W}^{-1/2}\mathbf{Z}_M$, $k = 1, \ldots, M$. Moreover, from expression (2.24) the sum $\mathbf{Z}'_M\mathcal{W}^{-1}\mathbf{Z}_M = \sum_{k=1}^{M}\boldsymbol{\Xi}_k\boldsymbol{\Xi}'_k$ is an approximation for the information matrix $\mathbf{I}(\boldsymbol{\Lambda})$. Equation (2.40) suggests a plot of the adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ with bands

$$\pm z_{\alpha/2}\, n^{-1/2}\sqrt{1 - \mathbf{a}'_m\boldsymbol{\Xi}'_k(\mathbf{Z}'_M\mathcal{W}^{-1}\mathbf{Z}_M)^{-1}\boldsymbol{\Xi}_k\mathbf{a}_m} \ , \quad 1 \le k \le M \ , \tag{2.41}$$

as a possible diagnostic check in *VARMA(p,q)* processes, where $z_{\alpha/2}$ is a suitable quantile of a $N(0,1)$ distribution. This extends a well-known tool in univariate *ARMA(p,q)* models based on the residual correlations $\widehat{r}_k$ of (1.5). See Brockwell and Davis (1991, chapter 9). As it will be established later in chapter 3, the quantities of (2.40) are close to one for $k$ large enough. The practical use of (2.41) requires replacing the population parameters $(\mathbf{\Phi}, \mathbf{\Theta}, \mathbf{\Sigma})$ by their ML estimators $(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}}, \widehat{\mathbf{\Sigma}})$.

Another application for goodness-of-fit purposes of the adjusted residual traces could be to consider, as mentioned in section 1.4.2, the process of expression (1.28),

$$\widehat{W}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n} \sum_{k=1}^{n-(P+1)} \frac{\operatorname{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}} \frac{\sin(k\pi u)}{k} \ .$$

The use of $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ is motivated by a model error process that is studied in the next section.

## 2.6 The error goodness-of-fit process

### 2.6.1 Introduction

In univariate *ARMA(p, q)* models, an alternative to the standard goodness-of-fit method of Box and Pierce (1970) is to work in the frequency domain (Priestley, 1981). One possibility is to consider the process of (1.25),

$$\widehat{W}_n(u) = \frac{\sqrt{2}}{\pi}\sqrt{n} \sum_{k=1}^{n-(P+1)} \widehat{r}_k \frac{\sin(k\pi u)}{k} \ .$$

This is just the sample version of $\{W_n(u) : 0 \leq u \leq 1\}$, where

$$W_n(u) = \frac{\sqrt{2}}{\pi}\sqrt{n} \sum_{k=1}^{n-1} r_k \frac{\sin(k\pi u)}{k} \tag{2.42}$$

depends on the sample correlations of the errors $\{\varepsilon_t : 1 \leq t \leq n\}$,

$$r_k = \frac{\sum_{t=1}^{n-k} \varepsilon_t \varepsilon_{t+k}}{\sum_{t=1}^{n} \varepsilon_t^2} \ , \quad 0 \leq k \leq n-1 \ ,$$

given in (1.6). As studied by Durlauf (1991, section 2) and Anderson (1993, section 2), under adequate regularity conditions the process of (2.42) converges weakly in $C[0,1]$ as $n \to \infty$ to the Brownian bridge $\{B(u) : 0 \leq u \leq 1\}$.

Consider the correlation matrices $\mathbf{R}_k = \mathbf{C}_k' \mathbf{C}_0^{-1}$ of the errors $\{\boldsymbol{\varepsilon}_t : 1 \leq t \leq n\}$ of model (1.1) (Chitturi, 1974), where from expression (1.19)

$$\mathbf{C}_k = \frac{1}{n} \sum_{t=1}^{n-k} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t+k}' \ , \quad 0 \leq k \leq n-1 \ .$$

For $m = 1$, the adjusted error traces

$$\mathrm{tr}(\mathbf{R}_k)/\sqrt{m}\ , \quad 1 \leq k \leq n-1\ , \tag{2.43}$$

coincide with the error autocorrelations $r_k$ of (1.6). Thus, a possible extension of (2.42) for *VARMA(p,q)* models is

$$W_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-1}\frac{\mathrm{tr}(\mathbf{R}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k}\ . \tag{2.44}$$

As its univariate counterpart of (2.42), the process $\{W_n^m(u) : 0 \leq u \leq 1\}$ of (2.44) is not feasible, because it depends on the unobservable quantities $\{\boldsymbol{\varepsilon}_t : 1 \leq t \leq n\}$. However, it will serve as a building block for a new procedure of goodness-of-fit in multivariate time series, presented in chapter 4.

The objective now is to establish that $\{W_n^m(u) : 0 \leq u \leq 1\}$ converges weakly, as $n \to \infty$, to the Brownian bridge. The corresponding derivations require a collection of auxiliary asymptotic results, that are collected in the next section.

### 2.6.2 Auxiliary asymptotic results

The first result refers to the convergence in distribution used in (2.29) $-$(2.30).

**Proposition 2.6.1** Suppose that the error vectors $\{\boldsymbol{\varepsilon}_t\}$ are i.i.d. with $\mathrm{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$; $\mathrm{Var}[\boldsymbol{\varepsilon}_t] = \boldsymbol{\Sigma} > 0$; and finite fourth order moments $\mathrm{E}[\|\boldsymbol{\varepsilon}_t\|^4] < +\infty$. Then, as $n \longrightarrow \infty$,

$$\sqrt{n}\begin{pmatrix} \mathrm{vec}(\mathbf{C}_1') \\ \mathrm{vec}(\mathbf{C}_2') \\ \vdots \\ \mathrm{vec}(\mathbf{C}_M') \end{pmatrix} \xrightarrow{D} \mathcal{W}^{1/2}\begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix}\ , \quad M \geq 1\ , \tag{2.45}$$

where the $\mathbf{V}_k$, $k = 1, \ldots, M$, are i.i.d. $N_{m^2}(\mathbf{0}, \mathbf{I}_{m^2})$; and $\mathcal{W} = \mathbf{I}_M \otimes \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}$.

**Proof.** The proof is given in Appendix 2.2. ∎

Proposition 2.6.1 generalizes well-known results relative to the convergence of the univariate statistics $\sqrt{n}\sum_{t=1}^{n-k}\varepsilon_t\varepsilon_{t+k}/n$, $1 \leq k \leq M$, to a $N_M(\mathbf{0}, \sigma^4\mathbf{I}_M)$ distribution (Brockwell and Davis, 1991, chapter 6). It is related to Chitturi (1976, theorem 1).

**Corollary 2.6.1** Put $\overline{\mathbf{R}}_k = \mathbf{C}_k'\mathbf{\Sigma}^{-1}$, $1 \le k \le n-1$. Under the same assumptions of Proposition 2.6.1, as $n \to \infty$

$$\sqrt{n}\,[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\overline{\mathbf{R}}_1) \\ \mathrm{tr}(\overline{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\overline{\mathbf{R}}_M) \end{pmatrix}] \xrightarrow{D} N_M(\mathbf{0}, \mathbf{I}_M)\ , \quad M \ge 1\ . \tag{2.46}$$

**Proof.** Convergence (2.46) follows from (2.45) using that

$$\sqrt{n}\,[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\overline{\mathbf{R}}_1) \\ \mathrm{tr}(\overline{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\overline{\mathbf{R}}_M) \end{pmatrix}] = (\mathbf{I}_k \otimes \mathbf{a}_m')\mathcal{W}^{-1/2}[\sqrt{n}\begin{pmatrix} \mathrm{vec}(\mathbf{C}_1') \\ \mathrm{vec}(\mathbf{C}_2') \\ \vdots \\ \mathrm{vec}(\mathbf{C}_M') \end{pmatrix}] \xrightarrow{D} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_M \end{pmatrix}\ , \quad M \ge 1\ ,$$

where the $v_k = \mathbf{a}_m'\mathbf{V}_k$ are i.i.d. $N(0,1)$ random variables, $k = 1, \ldots, M$. ∎

The main result of this section is given next.

**Proposition 2.6.2** Under the same assumptions of Proposition 2.6.1, as $n \to \infty$

$$\sqrt{n}\,[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\mathbf{R}_1) \\ \mathrm{tr}(\mathbf{R}_2) \\ \vdots \\ \mathrm{tr}(\mathbf{R}_M) \end{pmatrix}] \xrightarrow{D} N_M(\mathbf{0}, \mathbf{I}_M)\ , \quad M \ge 1\ . \tag{2.47}$$

**Proof.** Using the $Mm^2 \times Mm^2$ matrix $\mathbf{W} = \mathbf{I}_M \otimes \mathbf{C}_0 \otimes \mathbf{C}_0$, it can be written

$$\sqrt{n}\,[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\mathbf{R}_1) \\ \mathrm{tr}(\mathbf{R}_2) \\ \vdots \\ \mathrm{tr}(\mathbf{R}_M) \end{pmatrix}] = (\mathbf{I}_M \otimes \mathbf{a}_m')\mathbf{W}^{-1/2}[\sqrt{n}\begin{pmatrix} \mathrm{vec}(\mathbf{C}_1') \\ \mathrm{vec}(\mathbf{C}_2') \\ \vdots \\ \mathrm{vec}(\mathbf{C}_M') \end{pmatrix}]\ .$$

From a continuity argument similar to that in Brockwell and Davis (1991, Proposition 6.1.4) it can be checked that $\mathbf{W}^{-1/2} \xrightarrow{P} \mathcal{W}^{-1/2}$, where $\mathcal{W} = \mathbf{I}_M \otimes \mathbf{\Sigma} \otimes \mathbf{\Sigma} > \mathbf{0}$. Therefore, from Proposition 2.6.1 and Slutsky's theorem,

$$\sqrt{n}[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\mathbf{R}_1) \\ \mathrm{tr}(\mathbf{R}_2) \\ \vdots \\ \mathrm{tr}(\mathbf{R}_M) \end{pmatrix}] \xrightarrow{D} (\mathbf{I}_M \otimes \mathbf{a}_m')\mathcal{W}^{-1/2}\mathcal{W}^{1/2}\begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_M \end{pmatrix}\ ,$$

where the $v_k = \mathbf{a}_m'\mathbf{V}_k$, $k = 1, \ldots, M$, are as in Corollary 2.6.1. ∎

### 2.6.3 A result on the convergence of a stochastic process in $C[0,1]$

The following lemma will be used for establishing the convergence properties of the error process of (2.44). The proof is given in Ubierna and Velilla (2007, section 2).

**Lemma 2.6.1** Let $\{A_n(u) : 0 \le u \le 1\}$, $n = 1, 2, \ldots$, and $\{A(u) : 0 \le u \le 1\}$ be processes in $C[0,1]$. Consider a fixed integer $M_0 \ge 1$, and suppose that for each $M \ge M_0$ it can be written $A_n(u) = A_n^M(u) + R_n^M(u)$, and $A(u) = A^M(u) + R^M(u)$, $0 \le u \le 1$. Assume also the following three conditions for the processes $\{A_n^M(u) : 0 \le u \le 1\}$, $\{A^M(u) : 0 \le u \le 1\}$, $\{R_n^M(u) : 0 \le u \le 1\}$, and $\{R^M(u) : 0 \le u \le 1\}$:

**(C.1)** For each $M \ge M_0$, the finite-dimensional distributions of the sequence $\{A_n^M(u) : 0 \le u \le 1\}$ converge weakly, as $n \to \infty$, to those of $\{A^M(u) : 0 \le u \le 1\}$;

**(C.2)** For each $M \ge M_0$, the probability distributions of the sequence $\{A_n^M(u) : 0 \le u \le 1\}$ are tight;

**(C.3)** For each $\varepsilon > 0$

$$\limsup_M \,[\limsup_n \Pr( \sup_{0\le u\le 1} |R_n^M(u)| > \varepsilon)] = 0 \ ;$$

and $\limsup_M \Pr(\sup_{0\le u\le 1} |R^M(u)| > \varepsilon) = 0$.

Then, $\{A_n(u) : 0 \le u \le 1\}$ converges weakly in $C[0,1]$, as $n \to \infty$, to the process $\{A(u) : 0 \le u \le 1\}$.

### 2.6.4 Convergence of an auxiliary process

The convergence properties of (2.44) will be studied by analyzing first those of the auxiliary process $\{\overline{W}_n^m(u) : 0 \le u \le 1\}$, where

$$\overline{W}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-1} \frac{\mathrm{tr}(\overline{\mathbf{R}}_k)}{\sqrt{m}} \frac{\sin(k\pi u)}{k} \ , \tag{2.48}$$

and $\overline{\mathbf{R}}_k = \mathbf{C}_k'\mathbf{\Sigma}^{-1}$, $1 \le k \le n-1$, are the matrices considered in Corollary 2.6.1.

**Theorem 2.6.1** If the error vectors $\{\boldsymbol{\varepsilon}_t\}$ are i.i.d. with $\mathrm{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$; $\mathrm{Var}[\boldsymbol{\varepsilon}_t] = \mathbf{\Sigma} > 0$; and finite eighth order moments $\mathrm{E}[\|\boldsymbol{\varepsilon}_t\|^8] < +\infty$, then, as $n \to \infty$,

$$\overline{W}_n^m(u) \to_\omega B(u) \ .$$

**Proof.** Recall the standard Karhunen-Loève representation of the Brownian bridge

$$B(u) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{\infty} v_k \frac{\sin(k\pi u)}{k} \ , \quad 0 \le u \le 1 \ , \tag{2.49}$$

where $\{v_k : k \ge 1\}$ is a sequence of i.i.d. $N(0,1)$ random variables. See for example Ash and Gardner (1975, Chapter 1). In the light of expressions (2.48) and (2.49), the natural setting in Lemma 2.6.1 for establishing the convergence of $A_n(u) = \overline{W}_n^m(u)$ to $A(u) = B(u)$, is to select the integer $M_0 = 1$; and the processes

$$A_n^M(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{M} \frac{\mathrm{tr}(\overline{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k} \quad ; \quad R_n^M(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=M+1}^{n-1} \frac{\mathrm{tr}(\overline{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k} \ ;$$

$$A^M(u) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{M} v_k \frac{\sin(k\pi u)}{k} \quad ; \quad R^M(u) = \frac{\sqrt{2}}{\pi}\sum_{k=M+1}^{\infty} v_k \frac{\sin(k\pi u)}{k} \ .$$

In this manner, $A_n(u) = A_n^M(u) + R_n^M(u)$; and $A(u) = A^M(u) + R^M(u)$, $M \ge 1$. It is also convenient to consider the family of $M \times 1$ vectors, $M \ge 1$,

$$\boldsymbol{\alpha}_M(u) = \frac{\sqrt{2}}{\pi}\,[\sin(\pi u),\ \sin(2\pi u)/2,\ \ldots,\ \sin(M\pi u)/M]' \ , \quad 0 \le u \le 1 \ . \tag{2.50}$$

The rest of the proof is jut a a matter to check that, with the choices above, the conditions **(C.1)**–**(C.2)**–**(C.3)** of Lemma 2.6.1 hold.

**(C.1)** Pick $r \ge 1$, and select $u_1, u_2, \ldots, u_r$ in $[0,1]$. It can be written,

$$A_n^M(u) = \boldsymbol{\alpha}'_M(u)\sqrt{n}\ [\mathrm{tr}(\overline{\mathbf{R}}_1), \mathrm{tr}(\overline{\mathbf{R}}_2), \cdots, \mathrm{tr}(\overline{\mathbf{R}}_M)]'/\sqrt{m} \ .$$

Therefore, by Corollary 2.6.1 and the Cramér-Wold device,

$$\begin{pmatrix} A_n^M(u_1) \\ A_n^M(u_2) \\ \vdots \\ A_n^M(u_r) \end{pmatrix} = \mathbf{M}\sqrt{n}[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\overline{\mathbf{R}}_1) \\ \mathrm{tr}(\overline{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\overline{\mathbf{R}}_M) \end{pmatrix}] \xrightarrow{D} \mathbf{M}\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_M \end{pmatrix} = \begin{pmatrix} A^M(u_1) \\ A^M(u_2) \\ \vdots \\ A^M(u_r) \end{pmatrix} ,$$

where $\mathbf{M}$ is a $r \times M$ constant matrix whose $j$th row is $\boldsymbol{\alpha}'_M(u_j)$, $j = 1, \ldots, r$. ■

**(C.2)** For checking tightness of $\{A_n^M(u) : 0 \le u \le 1\}$, we use the necessary and sufficient conditions *(i)* and *(ii)* of Theorem 7.3 in Billingsley (1976, p. 82):

*(i)* Since $A_n^M(0) = 0$, the sequence of random variables $\{A_n^M(0)\}$, $n = 1, 2, \ldots$, is bounded in probability. Hence, condition *(i)* holds.

*(ii)* By the mean value theorem,

$$|\sin(k\pi u)/k - \sin(k\pi v)/k| \leq \pi\,|u - v| \ , \quad 0 \leq u,\, v \leq 1 \ . \tag{2.51}$$

From (2.51), $|A_n^M(u) - A_n^M(v)| \leq Z_n^M\,|u - v|$, where as a consequence of Corollary 2.6.1 the sequence $Z_n^M = (\sqrt{2}/\pi)\sqrt{n}\sum_{k=1}^M \left|\operatorname{tr}(\overline{\mathbf{R}}_k)/\sqrt{m}\right|$ is $O_P(1)$. Thus, given $\varepsilon$, $\eta > 0$, there exists $N(\eta) > 0$ such that $\Pr[Z_n^M > N(\eta)] \leq \eta$, $n \geq 1$. Hence, if $0 < \delta = \delta(\varepsilon, \eta) < \min[1, \varepsilon/N(\eta)]$, it follows that

$$\Pr[\sup_{|u-v|<\delta} \left|A_n^M(u) - A_n^M(v)\right| \geq \varepsilon] \leq \Pr[Z_n^M > N(\eta)] \leq \eta \ , \quad n \geq 1 \ .$$

As a conclusion, condition *(ii)* holds. ∎

**(C.3)** According to the results in Grenander and Rosenblatt (1957, Theorem 1, p. 188), the sequence of random variables $\{\sup_{0\leq u\leq 1}\left|R^M(u)\right| : M \geq 1\}$ converges to zero in probability as $M \to \infty$. This is enough to guarantee that the condition $\limsup_M \Pr(\sup_{0\leq u\leq 1}|R^M(u)| > \varepsilon) = 0$ holds for all $\varepsilon > 0$.

On the other hand, if $\{w_t\}$ are i.i.d. with $\mathrm{E}(w_t) = 0$; $\mathrm{var}(w_t) = 1$; and finite eighth order moment $E(|w_t|^8) < +\infty$, then putting $c_k = \sum_{t=1}^{n-k} w_t w_{t+k}/n$, $1 \leq k \leq n-1$,

$$\limsup_M\, [\limsup_n \Pr(\sup_{0\leq u\leq 1}\left|\frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=M+1}^{n-1} c_k\,\frac{\sin(k\pi u)}{k}\right| > \varepsilon)] = 0 \ . \tag{2.52}$$

Define now $\mathbf{w}_t = \mathbf{\Sigma}^{-1/2}\boldsymbol{\varepsilon}_t = (w_{t,1}, \ldots, w_{t,m})'$, $t \in \mathbb{Z}$. The $\{\mathbf{w}_t\}$ are a sequence of i.i.d random vectors with $E(\mathbf{w}_t) = \mathbf{0}$ and $\mathrm{Var}(\mathbf{w}_t) = \mathbf{I}_m$. Moreover, it is easy to see that, as an application of Hölder's inequality, $E(|w_{t,I}|^8) < +\infty$, $I = 1, \ldots, m$. In what comes next, it is convenient to consider the empirical covariances

$$c_{k,II} = \frac{1}{n}\sum_{t=1}^{n-k} w_{t,I}w_{t+k,I} \ , \quad 0 \leq k \leq n-1 \ , \tag{2.53}$$

of the univariate sequences $\{w_{t,I}\}$, $I = 1, \ldots, m$. It can be written $\operatorname{tr}(\overline{\mathbf{R}}_k) = \operatorname{tr}(\mathbf{C}_k'\mathbf{\Sigma}^{-1}) = \operatorname{tr}(\mathbf{\Sigma}^{-1/2}\mathbf{C}_k\mathbf{\Sigma}^{-1/2})$, where $\mathbf{C}_k = \sum_{t=1}^{n-k}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t+k}'/n$, $0 \leq k \leq n-1$. Thus, using definition (2.53) it follows that

$$\operatorname{tr}(\overline{\mathbf{R}}_k) = \frac{1}{n}\operatorname{tr}[\sum_{t=1}^{n-k}(\mathbf{\Sigma}^{-1/2}\boldsymbol{\varepsilon}_t)(\boldsymbol{\varepsilon}_{t+k}'\mathbf{\Sigma}^{-1/2})] = \frac{1}{n}\sum_{I=1}^{m}\sum_{t=1}^{n-k} w_{t,I}w_{t+k,I} = \sum_{I=1}^{m} c_{k,II} \ . \tag{2.54}$$

By (2.54), the inequality below holds

$$\sup_{0\leq u\leq 1}\left|\frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=M+1}^{n-1}\frac{\operatorname{tr}(\overline{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k}\right| \leq$$

$$\leq \frac{1}{\sqrt{m}} \sum_{I=1}^{m} \sup_{0 \leq u \leq 1} \left| \frac{\sqrt{2}}{\pi} \sqrt{n} \sum_{k=M+1}^{n-1} c_{k,II} \frac{\sin(k\pi u)}{k} \right| . \tag{2.55}$$

From the result (2.52) by Grenander and Rosenblatt (1957), it can be obtained that

$$\limsup_{M} [\limsup_{n} \Pr( \sup_{0 \leq u \leq 1} \left| \frac{\sqrt{2}}{\pi} \sqrt{n} \sum_{k=M+1}^{n-1} c_{k,II} \frac{\sin(k\pi u)}{k} \right| > \varepsilon)] = 0 \ , \quad I = 1, \ldots, m \ . \tag{2.56}$$

From expression (2.56) and inequality (2.55), it follows finally that the condition $\limsup_M [\limsup_n \Pr(\sup_{0 \leq u \leq 1} |R_n^M(u)| > \varepsilon)] = 0$ holds. ■

### 2.6.5 Convergence of the error process

For establishing the convergence properties of the error process $\{W_n^m(u) : 0 \leq u \leq 1\}$ of (2.44), consider the $m^2 \times 1$ random vectors

$$\mathbf{U}_n^m(u) = \frac{\sqrt{2}}{\pi} \sqrt{n} \sum_{k=1}^{n-1} (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2}) \mathrm{vec}(\mathbf{C}_k) \frac{\sin(k\pi u)}{k} \ , \quad 0 \leq u \leq 1 \ . \tag{2.57}$$

Since $\mathbf{R}_k = \mathbf{C}_k' \mathbf{C}_0^{-1}$, it can be written

$$W_n^m(u) = \mathbf{a}_{n,m}' \mathbf{U}_n^m(u) \quad , \quad \overline{W}_n^m(u) = \mathbf{a}_m' \mathbf{U}_n^m(u) \ , \quad 0 \leq u \leq 1 \ , \tag{2.58}$$

where $\mathbf{a}_{n,m} = (\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{1/2})(\mathbf{C}_0^{-1/2} \otimes \mathbf{C}_0^{-1/2}) \mathbf{a}_m$, and $\mathbf{a}_m = \mathrm{vec}(\mathbf{I}_m)/\sqrt{m}$ are $m^2 \times 1$ vectors with coordinates

$$\mathbf{a}_{n,m}(I,J) = [\mathrm{vec}(\mathbf{e}_I \mathbf{e}_J')]' \mathbf{a}_{n,m} \quad , \quad \mathbf{a}_m(I,J) = [\mathrm{vec}(\mathbf{e}_I \mathbf{e}_J')]' \mathbf{a}_m \ , \quad I, J = 1, \ldots, m \ ,$$

where $\mathbf{e}_I$ and $\mathbf{e}_J$ are the $I$th and $J$th canonical vectors of $\mathbb{R}^m$, $I, J = 1, \ldots, m$.

**Proposition 2.6.3** Under the same assumptions for the errors given in the statement of Theorem 2.6.1, it follows that

$$\sup_{0 \leq u \leq 1} \left| W_n^m(u) - \overline{W}_n^m(u) \right| = o_P(1) \ . \tag{2.59}$$

**Proof.** Using the representations of (2.58), it can be written $W_n^m(u) - \overline{W}_n^m(u) = (\mathbf{a}_n - \mathbf{a})' \mathbf{U}_n^m(u)$, $0 \leq u \leq 1$. Since $\mathbf{C}_0 \xrightarrow{P} \mathbf{\Sigma} > \mathbf{0}$, it follows that $\mathbf{a}_{n,m} \xrightarrow{P} (\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{1/2})(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2}) \mathbf{a}_m = (\mathbf{I}_m \otimes \mathbf{I}_m) \mathbf{a}_m = \mathbf{I}_{m^2} \mathbf{a}_m = \mathbf{a}_m$. To finish the proof of this proposition, it is then enough to check that all the coordinates $[\mathrm{vec}(\mathbf{e}_I \mathbf{e}_J')]' \mathbf{U}_n^m(u)$ of the $m^2 \times 1$ process $\{\mathbf{U}_n^m(u) : 0 \leq u \leq 1\}$ are bounded in probability.

From (2.57), for each $I, J = 1, \ldots, m$:

$$[\mathrm{vec}(\mathbf{e}_I\mathbf{e}_J')]'\mathbf{U}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-1}[\mathrm{vec}(\mathbf{e}_I\mathbf{e}_J')]'(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2})\mathrm{vec}(\mathbf{C}_k)\frac{\sin(k\pi u)}{k} \ . \tag{2.60}$$

Using the matrix properties of section 1.4.2, it can be written

$$[\mathrm{vec}(\mathbf{e}_I\mathbf{e}_J')]'(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2})\mathrm{vec}(\mathbf{C}_k) = \mathbf{e}_I'\boldsymbol{\Sigma}^{-1/2}\mathbf{C}_k\boldsymbol{\Sigma}^{-1/2}\mathbf{e}_J = \frac{1}{n}\sum_{t=1}^{n-k} w_{t,I}w_{t+k,I} \ , \tag{2.61}$$

where $\mathbf{w}_t = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\varepsilon}_t = (w_{t,1}, \ldots, w_{t,m})'$ are the normalized error vectors considered in the proof of Theorem 2.6.1. Therefore, the left-hand side of (2.60) is

$$[\mathrm{vec}(\mathbf{e}_I\mathbf{e}_J')]'\mathbf{U}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-1} c_{k,IJ}\frac{\sin(k\pi u)}{k} \ , \quad 0 \leq u \leq 1 \ , \tag{2.62}$$

where $c_{k,IJ} = \sum_{t=1}^{n-k} w_{t,I}w_{t+k,J}/n$, $0 \leq k \leq n-1$, is of the form (2.53). Proceeding then as in (2.52) by Grenander and Rosenblatt (1957), the probability bound $M_n = \max_{1\leq I,J\leq m}\sup_{0\leq u\leq 1}|[\mathrm{vec}(\mathbf{e}_I\mathbf{e}_J')]'\mathbf{U}_n^m(u)| = O_P(1)$ is obtained. Consequently,

$$\sup_{0\leq u\leq 1}|W_n^m(u) - \overline{W}_n^m(u)| = \sup_{0\leq u\leq 1}|(\mathbf{a}_n - \mathbf{a})'\mathbf{U}_n^m(u)| \leq$$
$$\leq m^2 \max_{1\leq I,J\leq m}|\mathbf{a}_n(I,J) - \mathbf{a}(I,J)|\, M_n = o_P(1) \ . \quad \blacksquare$$

Proposition 2.6.3 and Theorem 2.6.1 lead to the main result of this section, that characterizes the limit behavior of the error process $\{W_n^m(u) : 0 \leq u \leq 1\}$ of (2.44).

**Theorem 2.6.2** Under the same assumptions for the errors given in the statement of Theorem 2.6.1, as $n \to \infty$

$$W_n^m(u) \to_\omega B(u) \ .$$

**Proof.** By Theorem 7.1 in Billingsley (1976, p. 80), it suffices to prove that: (a) the finite dimensional distributions of $\{W_n^m(u) : 0 \leq u \leq 1\}$ converge weakly to those of $\{B(u) : 0 \leq u \leq 1\}$; and (b) the sequence of probability distributions of $\{W_n^m(u) : 0 \leq u \leq 1\}$ is tight.

Put $W_n^m(u) = \overline{W}_n^m(u) + D_n(u)$, where $D_n(u) = W_n^m(u) - \overline{W}_n^m$. $0 \leq u \leq 1$. For establishing (a), write

$$\begin{pmatrix} W_n(u_1) \\ W_n(u_2) \\ \vdots \\ W_n(u_r) \end{pmatrix} = \begin{pmatrix} \overline{W}_n(u_1) \\ \overline{W}_n(u_2) \\ \vdots \\ \overline{W}_n(u_r) \end{pmatrix} + \begin{pmatrix} D_n(u_1) \\ D_n(u_2) \\ \vdots \\ D_n(u_r) \end{pmatrix} \ . \tag{2.63}$$

From Lemma 2.6.1 and Theorem 2.6.1, the first summand at the left-hand side of (2.63) converges in distribution to the $r \times 1$ random vector $(B(u_1), \ldots, B(u_r))'$. From Proposition 2.6.3, the second summand of (2.63) is $o_P(1)$. Therefore, by Slutsky's theorem, $(W_n(u_1), \ldots, W_n(u_r))' \to_\omega (B(u_1), \ldots, B(u_r))'$.

For obtaining part (b), tightness of $\{W_n^m(u) : 0 \leq u \leq 1\}$, we use the necessary and sufficient conditions *(i)* and *(ii)* of Theorem 7.3 in Billingsley (1976, p. 82). Condition *(i)* holds trivially, since $W_n^m(0) = 0 = O_P(1)$. On the other hand, the inequality below holds,

$$\Pr(\sup_{|u-v|<\delta} |W_n^m(u) - W_n^m(v)| \geq \varepsilon) \leq$$

$$\leq \Pr(\sup_{|u-v|<\delta} \left|\overline{W}_n^m(u) - \overline{W}_n^m(v)\right| \geq \varepsilon/2) + \Pr(\sup_{0 \leq u \leq 1} |D_n(u)| \geq \varepsilon/4) \ . \tag{2.64}$$

From Lemma 2.6.1 and Theorem 2.6.1, the probability distributions of $\{\overline{W}_n^m(u) : 0 \leq u \leq 1\}$ are tight. Moreover, from Proposition 2.6.3, $\sup_{0 \leq u \leq 1} |D_n(u)| = o_P(1)$. Therefore, given $\varepsilon, \eta > 0$, there exist $0 < \delta < 1$ and $n_1$ such that the first summand of (2.64) is bounded above by $\eta/2$ for $n \geq n_1$. There exists also $n_2$ such that the second summand of (2.64) is below $\eta/2$ for $n \geq n_2$. In summary, condition *(ii)* of Theorem 7.3 in Billingsley (1976, p. 82) is satisfied for this $0 < \delta < 1$, and $n \geq \max(n_1, n_2)$. ■

## Appendix 2.1: Expressions of the blocks of the information matrix in section 2.3

Since the derivations of the expressions of all the blocks of the information matrix

$$\mathbf{I}(\mathbf{\Lambda}) = \begin{pmatrix} \mathbf{I}_{11}(\mathbf{\Lambda}) & \mathbf{I}_{12}(\mathbf{\Lambda}) \\ \mathbf{I}_{21}(\mathbf{\Lambda}) & \mathbf{I}_{22}(\mathbf{\Lambda}) \end{pmatrix}$$

are similar, only the case of $\mathbf{I}_{11}(\mathbf{\Lambda})$ is treated in detail. Recall the notation $\mathbf{\Lambda} = \mathrm{vec}(\mathbf{\Phi}, \mathbf{\Theta}) = [\lambda_1, \ldots, \lambda_{m^2(p+q)}]'$; $\mathbf{A}(\omega, \lambda_i) = \mathbf{k}^{-1}(\omega, \mathbf{\Lambda})\, [\partial \mathbf{k}(\omega, \mathbf{\Lambda})/\partial \lambda_i]$, $i = 1, \ldots, m^2(p+q)$, where $\mathbf{k}(\omega, \mathbf{\Lambda}) = \mathbf{\Phi}^{-1}(e^{i\omega})\mathbf{\Theta}(e^{i\omega})$; and $\mathbf{k}^{-1}(\omega, \mathbf{\Lambda}) = \mathbf{\Theta}^{-1}(e^{i\omega})\mathbf{\Phi}(e^{i\omega})$. Write also $\mathbf{\Phi}^{-1}(e^{i\omega})\mathbf{\Theta}(e^{i\omega}) = \sum_{j=0}^{\infty} \mathbf{\Omega}_j e^{ij\omega}$, and $\mathbf{\Theta}^{-1}(e^{i\omega}) = \sum_{j=0}^{\infty} \mathbf{L}_j e^{ij\omega}$.

Taking derivatives in the identity $\mathbf{\Phi}(e^{i\omega})\mathbf{k}(\omega, \mathbf{\Lambda}) = \mathbf{\Theta}(e^{i\omega})$, it follows that

$$\frac{\partial \mathbf{\Phi}(e^{i\omega})}{\partial \phi_{jk,r}} \mathbf{k}(\omega, \mathbf{\Lambda}) + \mathbf{\Phi}(e^{i\omega}) \frac{\partial \mathbf{k}(\omega, \mathbf{\Lambda})}{\partial \phi_{jk,r}} = \mathbf{0} \ , \tag{2.65}$$

where $\partial \mathbf{\Phi}(e^{i\omega})/\partial \phi_{jk,r} = -e^{ir\omega}\mathbf{e}_j \mathbf{e}_k'$. Therefore, from (2.65) it can be written

$$\begin{aligned} \mathbf{A}(\omega, \phi_{jk,r}) &= \mathbf{k}^{-1}(\omega, \mathbf{\Lambda})\, [\partial \mathbf{k}(\omega, \mathbf{\Lambda})/\partial \phi_{jk,r}] = \\ &= e^{ir\omega}\mathbf{k}^{-1}(\omega, \mathbf{\Lambda})\mathbf{\Phi}^{-1}(e^{i\omega})\mathbf{e}_j \mathbf{e}_k' \mathbf{k}(\omega, \mathbf{\Lambda}) = \\ &= e^{ir\omega}\mathbf{\Theta}^{-1}(e^{i\omega})\mathbf{e}_j \mathbf{e}_k' \mathbf{\Phi}^{-1}(e^{i\omega})\mathbf{\Theta}(e^{i\omega}) = \\ &= \sum_{u=0}^{\infty}\sum_{v=0}^{\infty} \mathbf{L}_u \mathbf{e}_j \mathbf{e}_k' \mathbf{\Omega}_v e^{i(r+u+v)\omega}. \end{aligned} \tag{2.66}$$

According to (2.66), it is obtained

$$\mathrm{tr}[\mathbf{A}(\omega, \phi_{jk,r})\mathbf{\Sigma}\mathbf{A}^*(\omega, \phi_{JK,R})\mathbf{\Sigma}^{-1}] = \tag{2.67}$$

$$\begin{aligned} &= \sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\sum_{U=0}^{\infty}\sum_{V=0}^{\infty} e^{i(r+u+v-R-U-V)\omega}\mathrm{tr}(\mathbf{L}_u \mathbf{e}_j \mathbf{e}_k' \mathbf{\Omega}_v \mathbf{\Sigma}\mathbf{\Omega}_V' \mathbf{e}_K \mathbf{e}_J' \mathbf{L}_U' \mathbf{\Sigma}^{-1}) = \\ &= \sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\sum_{U=0}^{\infty}\sum_{V=0}^{\infty} e^{i(r+u+v-R-U-V)w}\mathrm{tr}(\mathbf{\Omega}_V' \mathbf{e}_K \mathbf{e}_J' \mathbf{L}_U' \mathbf{\Sigma}^{-1}\mathbf{L}_u \mathbf{e}_j \mathbf{e}_k' \mathbf{\Omega}_v \mathbf{\Sigma}) = \\ &= \sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\sum_{U=0}^{\infty}\sum_{V=0}^{\infty} e^{i(r+u+v-R-U-V)w}[\mathrm{vec}(\mathbf{\Sigma}^{-1}\mathbf{L}_U \mathbf{e}_J \mathbf{e}_K' \mathbf{\Omega}_v)]'\mathrm{vec}(\mathbf{L}_u \mathbf{e}_j \mathbf{e}_k' \mathbf{\Omega}_v \mathbf{\Sigma}) \ . \end{aligned}$$

Notice that $[\mathrm{vec}(\mathbf{\Sigma}^{-1}\mathbf{L}_U \mathbf{e}_J \mathbf{e}_K' \mathbf{\Omega}_v)]'\mathrm{vec}(\mathbf{L}_u \mathbf{e}_j \mathbf{e}_k' \mathbf{\Omega}_v \mathbf{\Sigma}) = [\mathrm{vec}(\mathbf{e}_J \mathbf{e}_K')]'(\mathbf{\Omega}_V \otimes \mathbf{L}_U' \mathbf{\Sigma}^{-1})(\mathbf{\Sigma}\mathbf{\Omega}_v' \otimes \mathbf{L}_U)\mathrm{vec}(\mathbf{e}_j \mathbf{e}_k')$, where

$$(\mathbf{\Omega}_V \otimes \mathbf{L}_U' \mathbf{\Sigma}^{-1})(\mathbf{\Sigma}\mathbf{\Omega}_v' \otimes \mathbf{L}_u) \quad =$$

$$\begin{aligned} &= (\mathbf{\Omega}_V \mathbf{\Sigma} \otimes \mathbf{L}'_U)(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1})(\mathbf{\Sigma}\mathbf{\Omega}'_v \otimes \mathbf{L}_u) \quad = \\ &= (\mathbf{\Sigma}\mathbf{\Omega}'_V \otimes \mathbf{L}_U)'(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1})(\mathbf{\Sigma}\mathbf{\Omega}'_v \otimes \mathbf{L}_u) \ . \end{aligned}$$

It follows finally that $\mathrm{tr}[\mathbf{A}(\omega, \phi_{jk,r})\mathbf{\Sigma}\mathbf{A}^*(\omega, \phi_{JK,R})\mathbf{\Sigma}^{-1}]$ in (2.67) is obtained by left multiplying by $[\mathrm{vec}(\mathbf{e}_j\mathbf{e}'_k)]'$, and right multiplying by $\mathrm{vec}(\mathbf{e}_J\mathbf{e}'_K)$, the expression

$$\sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\sum_{U=0}^{\infty}\sum_{V=0}^{\infty} e^{i(r+u+v-R-U-V)w}(\mathbf{\Sigma}\mathbf{\Omega}'_v \otimes \mathbf{L}_u)'(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1})(\mathbf{\Sigma}\mathbf{\Omega}'_V \otimes \mathbf{L}_U) \ . \qquad (2.68)$$

Consider now the collection of matrices introduced in section 2.3,

$$\mathbf{G}_k = \sum_{j=0}^{k}(\mathbf{\Sigma}\mathbf{\Omega}'_j \otimes \mathbf{L}_{k-j}) \ , \quad 0 \le k \ ,$$

and put $\mathbf{G}_k = \mathbf{0}$, for $k < 0$. Introduce also the new index $L = u + v$. It is well-known that $\int_{-\pi}^{\pi} e^{ir\omega} d\omega = 2\pi$, for $r = 0$; and $\int_{-\pi}^{\pi} e^{ir\omega} d\omega = 0$, for $r = \pm 1, \pm 2, \ldots$. Hence, as far as taking an integral of the form (2.21) is concerned,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}[\mathbf{A}(\omega, \phi_{jk,r})\mathbf{\Sigma}\mathbf{A}^*(\omega, \phi_{JK,R})\mathbf{\Sigma}^{-1}]d\omega \ ,$$

the only set of indexes $(U, V)$ to be considered in expression (2.68) are those in which $L + r - R = U + V$. Accordingly, after multiplying in (2.68) on the left by $[\mathrm{vec}(\mathbf{e}_j\mathbf{e}'_k)]'$ and on the right by $\mathrm{vec}(\mathbf{e}_J\mathbf{e}'_K)$; integrating; and dividing by $2\pi$, the associated entry of the array $\mathbf{I}_{11}(\mathbf{\Lambda})$ is equal to

$$\begin{aligned} &[\mathrm{vec}(\mathbf{e}_j\mathbf{e}'_k)]'[\sum_{L=0}^{\infty} \mathbf{G}'_L(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1})\mathbf{G}_{L+(r-R)}]\mathrm{vec}(\mathbf{e}_J\mathbf{e}'_K) \quad = \\ &= [\mathrm{vec}(\mathbf{e}_j\mathbf{e}'_k)]'[\sum_{L=0}^{\infty} \mathbf{G}'_{L-r}(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1})\mathbf{G}_{L-R}]\mathrm{vec}(\mathbf{e}_J\mathbf{e}'_K) \ . \qquad (2.69) \end{aligned}$$

As a final conclusion, the $(r, R)$ block of $\mathbf{I}_{11}(\mathbf{\Delta})$ is the $m^2 \times m^2$ matrix $\sum_{L=0}^{\infty} \mathbf{G}'_{L-r}(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1})\mathbf{G}_{L-R}$, $r, R = 1, \ldots, p$.

An important particular case of the result above appears in a *VAR(p)* process. Then, $\mathbf{\Theta}(z) = \mathbf{I}_m$, and thus $\mathbf{G}_k = (\mathbf{\Sigma}\mathbf{H}'_k \otimes \mathbf{I}_m)$, where the $\{\mathbf{H}_k : k \ge 0\}$ are the coefficients of the series expansion $\mathbf{\Phi}^{-1}(z) = \sum_{k=0}^{\infty} \mathbf{H}_k\mathbf{z}^k$. Therefore, the $(r, R)$ block of the information matrix $\mathbf{I}[\text{vec}(\mathbf{\Phi})]$ is now the $m^2 \times m^2$ array

$$\begin{aligned}
& \sum_{k=0}^{\infty} \mathbf{G}'_{k-r}(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1})\mathbf{G}_{k-R} \quad = \\
= \quad & \sum_{k=0}^{\infty} (\mathbf{H}_{k-r} \otimes \mathbf{I}_m)(\mathbf{\Sigma} \otimes \mathbf{I}_m)(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1})(\mathbf{\Sigma} \otimes \mathbf{I}_m)(\mathbf{H}'_{k-R} \otimes \mathbf{I}_m) \quad = \\
= \quad & \sum_{k=0}^{\infty} (\mathbf{H}_{k-r} \otimes \mathbf{I}_m)(\mathbf{\Sigma} \otimes \mathbf{\Sigma}^{-1})(\mathbf{H}'_{k-R} \otimes \mathbf{I}_m) \quad = \\
= \quad & (\sum_{k=0}^{\infty} \mathbf{H}_{k-r}\mathbf{\Sigma}\mathbf{H}'_{k-R}) \otimes \mathbf{\Sigma}^{-1} \ .
\end{aligned} \tag{2.70}$$

Expression (2.70) above does not cancel the matrix $\mathbf{\Sigma}$. Also, the factor at the right of the Kronecker product symbol $\otimes$ depends only on the inverse $\mathbf{\Sigma}^{-1}$. ■

## Appendix 2.2: Proof of Proposition 2.6.1

Recall the definition $\mathbf{C}_k = n^{-1}\sum_{t=1}^{n-k}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+k}$ of (1.19). Then, it is easy to see that

$$\sqrt{n}\begin{pmatrix}\mathrm{vec}(\mathbf{C}_1)\\ \mathrm{vec}(\mathbf{C}_2)\\ \vdots\\ \mathrm{vec}(\mathbf{C}_M)\end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}\mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+1})\\ \mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+2})\\ \vdots\\ \mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+M})\end{pmatrix} + o_P(1)\ , \quad M \geq 1\ . \tag{2.71}$$

Given a collection of constant $m\times m$ matrices $\boldsymbol{\xi}_k$, $k=1,\ldots,M$, consider the sequence of random variables $\{X_t : t\in\mathbb{Z}\}$, where

$$X_t = \sum_{k=1}^{M}[\mathrm{vec}(\boldsymbol{\xi}_k)]'\mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+k}) = \sum_{k=1}^{M}\mathrm{tr}(\boldsymbol{\xi}'_k\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+k}) = \sum_{k=1}^{M}\boldsymbol{\varepsilon}'_t\boldsymbol{\xi}_k\boldsymbol{\varepsilon}_{t+k}\ . \tag{2.72}$$

Under the i.i.d. assumption on the $\{\boldsymbol{\varepsilon}_t\}$, the family $\{X_t : t\in\mathbb{Z}\}$ is strictly stationary. On the other hand, the sets $\{X_t : t \leq 0\}$ and $\{X_t : t \geq M+1\}$ are independent. Therefore, the sequence $\{X_t : t\in\mathbb{Z}\}$ is also $M$-dependent. Moreover,

$$\mathrm{E}(\boldsymbol{\varepsilon}'_t\boldsymbol{\xi}_k\boldsymbol{\varepsilon}_{t+k}) = \mathrm{E}[\mathrm{tr}(\boldsymbol{\xi}'_k\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+k})] = \mathrm{tr}[\boldsymbol{\xi}'_k E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+k})] = \mathrm{tr}[\boldsymbol{\xi}'_k\mathrm{Cov}(\boldsymbol{\varepsilon}_t\,,\boldsymbol{\varepsilon}'_{t+k})] = 0\ . \tag{2.73}$$

Hence, from representation(2.72) $\mathrm{E}(X_t)=0$. The covariance function is given by

$$\gamma(h) = \mathrm{E}(X_tX_{t+h}) =$$

$$= [\mathrm{vec}(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\ldots,\boldsymbol{\xi}_M)]'\mathrm{E}\left[\begin{pmatrix}\mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+1})\\ \mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+2})\\ \vdots\\ \mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+M})\end{pmatrix}\begin{pmatrix}\mathrm{vec}(\boldsymbol{\varepsilon}_{t+h}\boldsymbol{\varepsilon}'_{t+h+1})\\ \mathrm{vec}(\boldsymbol{\varepsilon}_{t+h}\boldsymbol{\varepsilon}'_{t+h+2})\\ \vdots\\ \mathrm{vec}(\boldsymbol{\varepsilon}_{t+h}\boldsymbol{\varepsilon}'_{t+h+M})\end{pmatrix}'\right]\mathrm{vec}(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\ldots,\boldsymbol{\xi}_M)\ . \tag{2.74}$$

Since $\mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+k}) = \boldsymbol{\varepsilon}_{t+k}\otimes\boldsymbol{\varepsilon}_t$, it follows that

$$\mathrm{E}\{\mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+k})[\mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+K})]'\} = \mathrm{E}[(\boldsymbol{\varepsilon}_{t+k}\otimes\boldsymbol{\varepsilon}_t)(\boldsymbol{\varepsilon}'_{t+K}\otimes\boldsymbol{\varepsilon}'_t)] = \mathrm{E}(\boldsymbol{\varepsilon}_{t+k}\boldsymbol{\varepsilon}'_{t+K}\otimes\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t)\ .$$

By the law of iterated expectations,

$$\mathrm{E}\{\mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+k})[\mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+K})]'\} = \mathrm{E}(\boldsymbol{\varepsilon}_{t+k}\boldsymbol{\varepsilon}'_{t+K}\otimes\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t) = \mathrm{E}[\mathrm{E}(\boldsymbol{\varepsilon}_{t+k}\boldsymbol{\varepsilon}'_{t+K}\otimes\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t|\boldsymbol{\varepsilon}_t)] =$$

$$= \mathrm{E}[\mathrm{Cov}(\boldsymbol{\varepsilon}_{t+k},\boldsymbol{\varepsilon}_{t+K})\otimes\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t] = \mathrm{Cov}(\boldsymbol{\varepsilon}_{t+k},\boldsymbol{\varepsilon}_{t+K})\otimes\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t) = \begin{cases}\mathbf{0} & ,\quad k\neq K\ .\\ \boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma} & ,\quad k=K\ .\end{cases}$$

Thus, from expression (2.74) it follows that $\gamma(h) = 0$ for $h \geq 1$, and

$$\gamma(0) = [\mathrm{vec}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_M)]' \mathcal{W} \mathrm{vec}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_M) \ ,$$

where $\mathcal{W} = \mathbf{I}_M \otimes \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}$. Using now theorem 6.4.2 in Brockwell and Davis (1991), the convergence result below holds:

$$\sqrt{n}\,[\mathrm{vec}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_M)]' \begin{pmatrix} \mathrm{vec}(\mathbf{C}_1) \\ \mathrm{vec}(\mathbf{C}_2) \\ \vdots \\ \mathrm{vec}(\mathbf{C}_M) \end{pmatrix} =$$

$$= \sqrt{n}\,(\frac{1}{n}\sum_{t=1}^{n} X_t) + o_P(1) \xrightarrow{D} [\mathrm{vec}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_M)]' \mathcal{W}^{1/2} \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix} \ . \tag{2.75}$$

Consider finally the $m^2 \times m^2$ commutation matrix $\mathbf{K}_{mm}$ of order $m$ (Lütkepohl, 2005, Sec. A.12.2); and the $Mm^2 \times Mm^2$ block-diagonal matrix $\mathbf{K} = \mathrm{diag}(\mathbf{K}_{mm}, \overset{(M)}{\ldots}, \mathbf{K}_{mm})$. Using the identity $\mathrm{vec}(\mathbf{C}'_k) = \mathbf{K}_{mm}\mathrm{vec}(\mathbf{C}_k)$ and the Cramér-Wold device, from (4.35) it is obtained that

$$\sqrt{n} \begin{pmatrix} \mathrm{vec}(\mathbf{C}'_1) \\ \mathrm{vec}(\mathbf{C}'_2) \\ \vdots \\ \mathrm{vec}(\mathbf{C}'_M) \end{pmatrix} = \mathbf{K}[\sqrt{n} \begin{pmatrix} \mathrm{vec}(\mathbf{C}_1) \\ \mathrm{vec}(\mathbf{C}_2) \\ \vdots \\ \mathrm{vec}(\mathbf{C}_M) \end{pmatrix}] \xrightarrow{D} \mathbf{K}\mathcal{W}^{1/2} \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix} \overset{D}{\equiv} \mathcal{W}^{1/2} \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix} \ . \tag{2.76}$$

The equivalence in distribution at the right-hand side of (2.76) follows from the identity $\mathbf{K}_{mm}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{K}_{mm} = \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}$ (Lütkepohl, 2005, Sec. A.12.2, Eq. (24)). Then, since $\mathbf{K}_{mm} = \mathbf{K}'_{mm}$, both $\mathbf{K}_{mm}(\boldsymbol{\Sigma}^{1/2} \otimes \boldsymbol{\Sigma}^{1/2})\mathbf{V}_k$ and $(\boldsymbol{\Sigma}^{1/2} \otimes \boldsymbol{\Sigma}^{1/2})\mathbf{V}_k$ have the same $N_{m^2}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})$ distribution, $k = 1, \ldots, M$. This finishes the proof of convergence (2.45) in Proposition 2.6.1. ■

# Chapter 3

# An initial goodness-of-fit process

**Summary.** This chapter analyzes the asymptotic properties of an empirical version of the error process $\{W_n^m(u) : 0 \le u \le 1\}$ defined in expression (2.44) of section 2.6. The idea is to replace the adjusted error traces $\mathrm{tr}(\mathbf{R}_k)/\sqrt{m}$, $1 \le k \le n-1$, of (2.43) by their residual counterparts $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$, $1 \le k \le n-(P+1)$, of (1.29). This leads to considering the residual process $\{\widehat{W}_n^m(u) : 0 \le u \le 1\}$ that appears in (1.28). Section 3.1 introduces the notation, and gives some background and motivation. Section 3.2 conjectures a possible form for the covariance function of the limit process. This can be written as a function of the $m^2 \times m^2(p+q)$ row blocks $\Xi_k'$ of the matrix $\mathcal{W}^{-1/2}\mathbf{Z}_M$, that are defined in (2.23). Section 3.3 contains the formal limit result. Section 3.4 gives some final comments.

## 3.1 Introduction

The error vectors $\{\boldsymbol{\varepsilon}_t : 1 \le t \le n\}$ are not observable. Therefore, they must be estimated by the residuals of (2.4), $\{\widehat{\boldsymbol{\varepsilon}}_t : P < t \le n\}$. Hence, a natural approach for goodness-of-fit in *VARMA(p,q)* models is to replace in expression (2.44),

$$W_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-1}\frac{\mathrm{tr}(\mathbf{R}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k}\ ,$$

the adjusted error traces $\mathrm{tr}(\mathbf{R}_k)/\sqrt{m}$, $1 \le k \le n-1$, of (2.43) by their residual counterparts of (1.29). This leads to considering the process $\{\widehat{W}_n^m(u) : 0 \le u \le 1\}$ of (1.28), where

$$\widehat{W}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-(P+1)}\frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k}$$

depends on the $m \times m$ residual autocorrelation matrices $\widehat{\mathbf{R}}_k = \widehat{\mathbf{C}}_k'\widehat{\boldsymbol{\Sigma}}^{-1}$, $1 \le k \le n-(P+1)$, of expression (1.21) (Chitturi, 1974).

For $m = 1$, the adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ coincide with the residual correlations $\widehat{r}_k$ of (1.5). Thus, $\{\widehat{W}_n^m(u) : 0 \le u \le 1\}$ generalizes the process considered by Ubierna and Velilla (2007, section 1), $\{\widehat{W}_n(u) : 0 \le u \le 1\}$, where

$$\widehat{W}_n(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-(P+1)}\widehat{r}_k\,\frac{\sin(k\pi u)}{k}$$

is as given in (1.25). Put $f(\omega) = (2\pi)^{-1} \mid \theta(e^{-i\omega}) \mid^2 / \mid \phi(e^{-i\omega}) \mid^2$ , $-\pi \le \omega \le \pi$, for the standardized spectral density of a univariate *ARMA(p,q)* process $\phi(B)(X_t-\mu) =$

$\theta(B)\varepsilon_t$. Let $\mathbf{I}(\boldsymbol{\lambda})$ denote the $(p+q)\times(p+q)$ information matrix for the $(p+q)\times 1$ vector of parameters $\boldsymbol{\lambda} = (\phi_1,\ldots,\phi_p;\theta_1,\ldots,\theta_q)'$. Ubierna and Velilla (2007, Theorem 2.1, p. 2905) establish that, as $n\to\infty$, $\widehat{W}_n(u) \to_w G(u)$, where $\{G(u) : 0\le u\le 1\}$ is a zero mean Gaussian process with covariance function

$$\gamma(u,v) = [\min(u,v) - uv] - \frac{1}{2\pi^2}\,\mathbf{g}(\pi u)'\mathbf{I}^{-1}(\boldsymbol{\lambda})\mathbf{g}(\pi v)\ , \qquad 0\le u,v\le 1\ , \tag{3.1}$$

and $\mathbf{g}(\pi u) = \int_0^{\pi u}[\,\partial \log f(\omega)/\partial\boldsymbol{\lambda}\,]\,d\omega$ is a $(p+q)\times 1$ vector. The first summand in (3.1) is the covariance function of the Brownian bridge $\{B(u) : 0\le u\le 1\}$; the second is a positive parametric quadratic form that does not depend on the scale parameter $\sigma^2$.

In the $m>1$ case, it can be motivated that the residual process $\{\widehat{W}_n^m(u) : 0\le u\le 1\}$ of (1.28) converges weekly to a zero mean Gaussian process $\{G^m(u) : 0\le u\le 1\}$ with covariance function of the form

$$\gamma^m(u,v) = [\min(u,v) - uv] - \frac{1}{2\pi^2 m}\,\mathbf{g}^m(\pi u)'\mathbf{I}^{-1}(\boldsymbol{\Lambda})\mathbf{g}^m(\pi v)\ , \qquad 0\le u,v\le 1\ , \tag{3.2}$$

where $\mathbf{I}(\boldsymbol{\Lambda})$ is the $m^2(p+q)\times m^2(p+q)$ information matrix for $\boldsymbol{\Lambda} = \mathrm{vec}(\boldsymbol{\Phi},\boldsymbol{\Theta}) = [\lambda_1,\ldots,\lambda_{m^2(p+q)}]'$; and $\mathbf{g}^m(\pi u)$ is a $m^2(p+q)\times 1$ vector of coordinates

$$\int_0^{\pi u} \mathrm{tr}\,[\mathbf{f}^{-1}(\omega,\boldsymbol{\Lambda})\,\partial\mathbf{f}(\omega,\boldsymbol{\Lambda})/\partial\lambda_i]\,d\omega\ , \quad i = 1,\ldots,m^2(p+q)\ , \tag{3.3}$$

that depend on the $m\times m$ spectral density matrix of the *VARMA(p,q)* model of (1.1),

$$\mathbf{f}(\omega,\boldsymbol{\Lambda}) = \frac{1}{2\pi}\,\boldsymbol{\Phi}^{-1}(e^{i\omega})\boldsymbol{\Theta}(e^{i\omega})\boldsymbol{\Sigma}\boldsymbol{\Theta}'(e^{-i\omega})\boldsymbol{\Phi}'^{-1}(e^{-i\omega})\ , \qquad -\pi\le\omega\le\pi\ . \tag{3.4}$$

The structure of (3.2) is similar to that of the univariate case in (3.1). However, by the results of section 2.3 the $m^2(p+q)\times m^2(p+q)$ matrix $\mathbf{I}(\boldsymbol{\Lambda})$ is not scale free, because it depends on the covariance matrix $\boldsymbol{\Sigma}$ of the errors $\{\boldsymbol{\varepsilon}_t\}$. Convergence of the residual process $\{\widehat{W}_n^m(u) : 0\le u\le 1\}$ to $\{G^m(u) : 0\le u\le 1\}$ requires some auxiliary results, that are studied next.

## 3.2 The explicit form of the limit covariance function

### 3.2.1 Motivation

Consider a sequence $\{\mathbf{V}_k : k \geq 1\}$ of i.i.d. $N_{m^2}(\mathbf{0}, \mathbf{I}_{m^2})$ random vectors. By expression (2.38) it follows that

$$\sqrt{n}\,[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{R}}_1) \\ \mathrm{tr}(\widehat{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{R}}_M) \end{pmatrix}] \stackrel{D}{\cong} (\mathbf{I}_M \otimes \mathbf{a}_m')(\mathbf{I}_{Mm^2} - \mathbf{P}_M)\begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix} ,$$

where $\mathbf{P}_M = \mathcal{W}^{-1/2}\mathbf{Z}_M(\mathbf{Z}_M'\mathcal{W}^{-1}\mathbf{Z}_M)^{-1}\mathbf{Z}_M'\mathcal{W}^{-1/2}$ is the $Mm^2 \times Mm^2$ orthogonal projection matrix onto the manifold spanned by the columns of $\mathcal{W}^{-1/2}\mathbf{Z}_M$, where $\mathbf{Z}_M = (\mathbf{X}_M, \mathbf{Y}_M)$ is as defined by expressions (2.19) and (2.20) of section 2.3.

Recall the notation of (2.23),

$$\mathbf{\Xi}_k' = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})(\mathbf{G}_{k-1}, \ldots, \mathbf{G}_{k-p}\,; \mathbf{F}_{k-1}, \ldots, \mathbf{F}_{k-q}) ,$$

for the $k$th $m^2 \times m^2(p+q)$ row block of the matrix $\mathcal{W}^{-1/2}\mathbf{Z}_M$, $k = 1, \ldots, M$, where $\mathcal{W} = \mathbf{I}_M \otimes \mathbf{\Sigma} \otimes \mathbf{\Sigma}$. By definition,

$$\mathbf{P}_M = \begin{pmatrix} \mathbf{P}_{11,M} & \mathbf{P}_{12,M} & \cdots & \mathbf{P}_{1M,M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{M1,M} & \mathbf{P}_{M2,M} & \cdots & \mathbf{P}_{MM,M} \end{pmatrix} , \tag{3.5}$$

where $\mathbf{P}_{jk,M} = \mathbf{\Xi}_j'(\mathbf{Z}_M'\mathcal{W}^{-1}\mathbf{Z}_M)^{-1}\mathbf{\Xi}_k$ is a $m^2 \times m^2$ matrix, $j, k = 1, \ldots, M$. Using the limit result (2.24) of section 2.3, it follows that as $M \to \infty$

$$\mathbf{P}_{jk,M} \to \mathbf{P}_{jk} = \mathbf{\Xi}_j'\mathbf{I}^{-1}(\mathbf{\Lambda})\mathbf{\Xi}_k , \quad j, k \geq 1 . \tag{3.6}$$

Since $\mathbf{P}_M$ is symmetric and idempotent, it is easily obtained from representation (3.5) above that $\mathbf{P}_{jk,M} = \sum_{s=1}^{M} \mathbf{P}_{js,M}\mathbf{P}_{sk,M}$, $j, k = 1, \ldots, M$. Hence, after taking the limit in the last expression as $M \to \infty$ using (3.6), the identity

$$\mathbf{P}_{jk} = \sum_{s=1}^{\infty} \mathbf{P}_{js}\mathbf{P}_{sk} , \qquad j, k \geq 1 , \tag{3.7}$$

holds. The justification of (3.7) is given in Appendix 3.1.

Consider now, as in the proof of Theorem 2.6.1, the $M \times 1$ vectors of (2.50),

$$\boldsymbol{\alpha}_M(u) = \frac{\sqrt{2}}{\pi}\,[\sin(\pi u),\ \sin(2\pi u)/2,\ \ldots,\ \sin(M\pi u)/M]'\ , \quad 0 \le u \le 1\ ,$$

where $M \ge 1$. For $n$ and $M$ large enough, it follows from (1.28) and (2.38) that

$$\widehat{W}_n^m(u) \overset{D}{\cong} \boldsymbol{\alpha}'_M(u)(\mathbf{I}_M \otimes \mathbf{a}'_m)(\mathbf{I}_{Mm^2} - \mathbf{P}_M)\begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix}\ , \quad 0 \le u \le 1\ . \tag{3.8}$$

Using (3.8), it can be written to some approximation

$$\mathrm{Cov}[\widehat{W}_n^m(u), \widehat{W}_n^m(v)] \cong$$

$$\cong \boldsymbol{\alpha}'_M(u)(\mathbf{I}_M \otimes \mathbf{a}'_m)(\mathbf{I}_{Mm^2} - \mathbf{P}_M)\mathbf{I}_{Mm^2}(\mathbf{I}_{Mm^2} - \mathbf{P}_M)(\mathbf{I}_M \otimes \mathbf{a}_m)\boldsymbol{\alpha}_M(v)\ . \tag{3.9}$$

Taking into account that $\mathbf{I}_{Mm^2} - \mathbf{P}_M$ is an orthogonal projection matrix, the right hand side of expression (3.9) can be expressed as the difference

$$\boldsymbol{\alpha}'_M(u)\boldsymbol{\alpha}_M(v) - \boldsymbol{\alpha}'_M(u)(\mathbf{I}_M \otimes \mathbf{a}'_m)\mathbf{P}_M(\mathbf{I}_M \otimes \mathbf{a}_m)\boldsymbol{\alpha}_M(v)\ . \tag{3.10}$$

From definition (2.50), the first summand of (3.10) can be written in the form $(2/\pi^2)\sum_{k=1}^M \sin(k\pi u)\sin(k\pi v)/k^2$. On the other hand, from (3.5) it follows that

$$(\mathbf{I}_M \otimes \mathbf{a}'_m)\mathbf{P}_M(\mathbf{I}_M \otimes \mathbf{a}_m) = \begin{pmatrix} \mathbf{a}'_m\mathbf{P}_{11,M}\mathbf{a}_m & \mathbf{a}'_m\mathbf{P}_{12,M}\mathbf{a}_m & \cdots & \mathbf{a}'_m\mathbf{P}_{1M,M}\mathbf{a}_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}'_m\mathbf{P}_{M1,M}\mathbf{a}_m & \mathbf{a}'_m\mathbf{P}_{M2,M}\mathbf{a}_m & \cdots & \mathbf{a}'_m\mathbf{P}_{MM,M}\mathbf{a}_m \end{pmatrix}\ .$$

Therefore, the second summand in (3.10) is $(2/\pi^2)\sum_{j=1}^M\sum_{k=1}^M(\mathbf{a}'_m\mathbf{P}_{jk,M}\mathbf{a}_m)\sin(j\pi u)$ $\sin(k\pi v)/jk$. Consequently, after taking the limit in (3.10) as $M \to \infty$ and using (3.6) it can be conjectured that the limit covariance function $\gamma^m(u,v)$, $0 \le u, v \le 1$, of the residual process $\{\widehat{W}_n^m(u) : 0 \le u \le 1\}$ is

$$\gamma^m(u,v) = \frac{2}{\pi^2}[\sum_{k=1}^\infty \frac{\sin(k\pi u)\sin(k\pi v)}{k^2} - \sum_{j=1}^\infty\sum_{k=1}^\infty(\mathbf{a}'_m\mathbf{P}_{jk}\mathbf{a}_m)\frac{\sin(j\pi u)}{j}\frac{\sin(k\pi v)}{k}]\ . \tag{3.11}$$

### 3.2.2 Structure

The first summand of expression (3.11),

$$\frac{2}{\pi^2}\sum_{k=1}^\infty \frac{\sin(k\pi u)\sin(k\pi v)}{k^2} = \min(u,v) - uv\ , \qquad 0 \le u, v \le 1\ ,$$

is the well-known covariance function of the Brownian bridge $\{B(u) : 0 \leq u \leq 1\}$. See e.g. Ash and Gardner (1975). Consequently, the first summand of (3.11) coincides with that in (3.2). To establish that the second component of expression (3.11) can be also written as that in (3.2), it suffices to check the identity

$$\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}(\mathbf{a}'\mathbf{P}_{jk}\mathbf{a})\frac{\sin(j\pi u)}{j}\frac{\sin(k\pi v)}{k} = \frac{1}{4m}\,\mathbf{g}^m(\pi u)'\mathbf{I}^{-1}(\mathbf{\Lambda})\mathbf{g}^m(\pi v)\ , \qquad 0 \leq u, v \leq 1\ , \tag{3.12}$$

where $\mathbf{g}^m(\pi u)$ is a $m^2(p+q) \times 1$ vector whose coordinates (3.3) depend on the partial derivatives of the $m \times m$ spectral density matrix $\mathbf{f}(\omega, \mathbf{\Lambda})$ of (3.4). Derivations, that are slightly cumbersome, are given in Appendix 3.2.

In view of the above, it can be written

$$\gamma^m(u, v) = [\min(u, v) - uv] - \frac{2}{\pi^2}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}(\mathbf{a}_m'\mathbf{P}_{jk}\mathbf{a}_m)\frac{\sin(j\pi u)}{j}\frac{\sin(k\pi v)}{k}\ . \tag{3.13}$$

By Appendix 3.1, the Euclidean norm of the matrices at the right-hand side of (3.6) is such that $\|\mathbf{P}_{jk}\| \leq ab^{j+k}$, $j, k \geq 1$, where $a > 0$, and $0 < b < 1$. Thus, the quantities $\mathbf{a}_m'\mathbf{P}_{jk}\mathbf{a}_m$ will be close to zero for $j, k$ large enough. As a consequence, the covariance function of (3.13) behaves in practice as that of the Brownian bridge corrected by a finite linear combination of the form $(2/\pi^2)\sum_{j=1}^{M}\sum_{k=1}^{M}(\mathbf{a}_m'\mathbf{P}_{jk}\mathbf{a}_m)\sin(j\pi u)\sin(k\pi v)/jk$, where $M$ is a suitable integer number. This property will be illustrated later in the examples of chapter 5.

## 3.3 A representation for the limit process

This section suggests a weak limit for the residual process of (1.28), that exploits the structure of (3.11). Consider the $m^2 \times m^2$ matrix function

$$\mathbf{p}_k(u) = \sum_{j=1}^{\infty}(\delta_{jk}\mathbf{I}_{m^2} - \mathbf{P}_{jk})\,\frac{\sin(j\pi u)}{j}\ , \quad 0 \leq u \leq 1\ , \tag{3.14}$$

where $\delta_{jk}$ is Dirac's delta. Given a sequence $\{\mathbf{V}_k : k \geq 1\}$ of i.i.d. $N_{m^2}(\mathbf{0}, \mathbf{I}_{m^2})$ random vectors, define the centered Gaussian process $\{G^m(u) : 0 \leq u \leq 1\}$, where

$$G^m(u) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{\infty}\mathbf{a}_m'\mathbf{p}_k(u)\mathbf{V}_k\ , \quad 0 \leq u \leq 1\ . \tag{3.15}$$

**Proposition 3.3.1** The covariance function of the process $\{G^m(u) : 0 \leq u \leq 1\}$ of (3.15) coincides with that given in (3.11).

**Proof.** From (3.15) it follows that

$$\mathrm{cov}[G^m(u), G^m(v)] = \frac{2}{\pi^2} \sum_{k=1}^{\infty} \mathbf{a}'_m \mathbf{p}_k(u) \mathbf{p}'_k(v) \mathbf{a}_m \ , \qquad 0 \leq u, v \leq 1 \ .$$

Using definition (3.14), it can be written

$$\sum_{k=1}^{\infty} \mathbf{p}_k(u)\mathbf{p}'_k(v) = \sum_{r=1}^{\infty}\sum_{s=1}^{\infty}[\sum_{k=1}^{\infty}(\delta_{rk}\mathbf{I}_{m^2} - \mathbf{P}_{rk})(\delta_{sk}\mathbf{I}_{m^2} - \mathbf{P}'_{sk})]\frac{\sin(r\pi u)}{r}\frac{\sin(s\pi v)}{s} \ .$$

On the other hand, from expressions (3.6) and (3.7), it follows that $\mathbf{P}'_{sr} = \mathbf{P}_{rs}$, and $\mathbf{P}_{rs} = \sum_{k=1}^{\infty} \mathbf{P}_{rk}\mathbf{P}_{ks}$, $r, s \geq 1$. Therefore,

$$\sum_{k=1}^{\infty}(\delta_{rk}\mathbf{I}_{m^2} - \mathbf{P}_{rk})(\delta_{sk}\mathbf{I}_{m^2} - \mathbf{P}'_{sk}) =$$

$$= \delta_{rs}\mathbf{I}_{m^2} - \mathbf{P}'_{sr} - \mathbf{P}_{rs} + \sum_{k=1}^{\infty}\mathbf{P}_{rk}\mathbf{P}'_{sk} = \delta_{rs}\mathbf{I}_{m^2} - \mathbf{P}_{rs} \ .$$

As a consequence, for $0 \leq u, v \leq 1$,

$$\mathrm{cov}[G^m(u), G^m(v)] = \frac{2}{\pi^2}\,\mathbf{a}'_m\,[\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}(\delta_{rs}\mathbf{I}_{m^2} - \mathbf{P}_{rs})\frac{\sin(r\pi u)}{r}\frac{\sin(s\pi v)}{s}]\,\mathbf{a}_m =$$

$$= \frac{2}{\pi^2}\,[\sum_{r=1}^{\infty}\frac{\sin(r\pi u)\sin(r\pi v)}{r^2} - \sum_{r=1}^{\infty}\sum_{s=1}^{\infty}(\mathbf{a}'_m\mathbf{P}_{rs}\mathbf{a}_m)\frac{\sin(r\pi u)}{r}\frac{\sin(s\pi v)}{s}] = \gamma^m(u,v) \ . \qquad \blacksquare$$

From Proposition 3.3.1 above and the arguments of Section 3.2, the process $\{G^m(u) : 0 \leq u \leq 1\}$ of (3.15) appears as a natural limit candidate for the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (1.28). Specific details are given next.

## 3.4 Weak convergence of the residual process

This section formalizes the weak convergence of the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (1.28) to the Gaussian random function of (3.15).

**Theorem 3.4.1** Under the same assumptions for the errors of model (1.1) than those given in Theorem 2.6.2, as $n \to \infty$

$$\widehat{W}_n^m(u) \to_\omega G^m(u) \ . \qquad (3.16)$$

**Proof.** The proof is again based on applying Lemma 2.6.1. The first step is to choose the adequate decompositions

$$\widehat{W}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k} = A_n^M(u) + R_n^M(u) \ ;$$

and

$$G^m(u) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{\infty} \mathbf{a}'_m \mathbf{p}_k(u)\mathbf{V}_k = A^M(u) + R^M(u) \ ,$$

for $M \geq M_0$, where $M_0$ is a properly selected fixed integer number.

**First part.** For choosing $A_n^M(u)$, write

$$\frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{M} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k} = \boldsymbol{\alpha}'_M(u)(\mathbf{I}_M \otimes \mathbf{a}'_m)\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}\sqrt{n}\mathcal{W}^{-1/2}\widehat{\mathbf{G}}_M \ , \quad (3.17)$$

where $\widehat{\mathbf{G}}_M = \{[\mathrm{vec}(\widehat{\mathbf{C}}'_1)]', \ldots, [\mathrm{vec}(\widehat{\mathbf{C}}'_M)]'\}'$ is a $Mm^2 \times 1$ vector, and $\boldsymbol{\alpha}_M(u)$ is the $M \times 1$ vector of (2.50). The behavior of the term $\boldsymbol{\alpha}'_M(u)(\mathbf{I}_M \otimes \mathbf{a}'_m)\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}$ in (3.17) is easy to characterize. Therefore, the choice of $A_n^M(u)$ depends on an analysis of the structure of the factor $\sqrt{n}\,\mathcal{W}^{-1/2}\widehat{\mathbf{G}}_M$.

Consider the matrices of $Mm^2 \times Mm^2$, $\mathbf{\Pi}_M = (\mathbf{P}_{jk}\colon 1 \leq j,k \leq M)$ and $\mathbf{P}_M = (\mathbf{P}_{jk,M}\colon 1 \leq j,k \leq M)$. Write

$$\sqrt{n}\,\mathcal{W}^{-1/2}\widehat{\mathbf{G}}_M = \mathbf{P}_M\sqrt{n}\,\mathcal{W}^{-1/2}\widehat{\mathbf{G}}_M + (\mathbf{I}_{Mm^2} - \mathbf{P}_M)\sqrt{n}\,\mathcal{W}^{-1/2}\widehat{\mathbf{G}}_M \ . \quad (3.18)$$

As a result of the orthogonality conditions in (2.28), the first summand at the right-hand side of (3.18) goes to $\mathbf{0}$ in probability. On the other hand, using expression (2.27) it follows that

$$\mathcal{W}^{-1/2}\widehat{\mathbf{G}}_M = \mathcal{W}^{-1/2}\mathbf{G}_M - \mathcal{W}^{-1/2}\mathbf{Z}_M\mathrm{vec}[(\widehat{\mathbf{\Phi}},\widehat{\mathbf{\Theta}}) - (\mathbf{\Phi},\mathbf{\Theta})] + O_P(\frac{1}{n}) \ , \quad (3.19)$$

where $\mathbf{G}_M = \{[\mathrm{vec}(\mathbf{C}'_1)]', \ldots, [\mathrm{vec}(\mathbf{C}'_M)]'\}'$ is a $Mm^2 \times 1$ vector. For $M \geq M_0 = p+q$, the orthogonal projection matrix $\mathbf{P}_M$ satisfies $(\mathbf{I}_{Mm^2} - \mathbf{P}_M)\mathcal{W}^{-1/2}\mathbf{Z}_M = \mathbf{0}$. Therefore, from (3.19) the second summand in (3.18) behaves as

$$(\mathbf{I}_{Mm^2} - \mathbf{\Pi}_M)\sqrt{n}\,\mathcal{W}^{-1/2}\mathbf{G}_M + (\mathbf{\Pi}_M - \mathbf{P}_M)\sqrt{n}\,\mathcal{W}^{-1/2}\mathbf{G}_M + O_P(\frac{1}{\sqrt{n}}) \ . \quad (3.20)$$

According to Appendix 3.1, $\mathbf{\Pi}_M - \mathbf{P}_M$ is close to $\mathbf{0}$ for $M$ large enough. Thus, by Proposition 2.6.1 the second summand in (3.20) is also ignorable in probability. Consequently, the natural choice for the term $A_n^M(u)$ is

$$A_n^M(u) = \boldsymbol{\alpha}'_M(u)(\mathbf{I}_M \otimes \mathbf{a}'_m)\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}(\mathbf{I}_{Mm^2} - \mathbf{\Pi}_M)\sqrt{n}\,\mathcal{W}^{-1/2}\mathbf{G}_M \ . \quad (3.21)$$

Therefore, going back to (3.17) and collecting terms together, it also follows that

$$R_n^M(u) = P_n^M(u) + Q_n^M(u) + S_n^M(u) \ , \quad 0 \le u \le 1 \ , \tag{3.22}$$

where

$$P_n^M(u) = \boldsymbol{\alpha}'_M(u)(\mathbf{I}_M \otimes \mathbf{a}'_m)\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}(\mathbf{\Pi}_M - \mathbf{P}_M)\sqrt{n}\,\mathcal{W}^{-1/2}\mathbf{G}_M \ ;$$

$$Q_n^M(u) = \boldsymbol{\alpha}'_M(u)(\mathbf{I}_M \otimes \mathbf{a}'_m)\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}\mathbf{P}_M\sqrt{n}\,\mathcal{W}^{-1/2}\widehat{\mathbf{G}}_M \ ;$$

and

$$S_n^M(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=M+1}^{n-(P+1)}\frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k} \ .$$

**Second part.** For choosing $A^M(u)$ in accordance with the structure of $A_n^M(u)$ in (3.21), recall that from (3.14)

$$\frac{\sqrt{2}}{\pi}\sum_{k=1}^{M}\mathbf{a}'_m\mathbf{p}_k(u)\mathbf{V}_k =$$

$$= \frac{\sqrt{2}}{\pi}\sum_{k=1}^{M}\mathbf{a}'_m[\sum_{j=1}^{M}(\delta_{jk}\mathbf{I}_{m^2} - \mathbf{P}_{jk})\frac{\sin(j\pi u)}{j} + \sum_{j=M+1}^{\infty}(\delta_{jk}\mathbf{I}_{m^2} - \mathbf{P}_{jk})\frac{\sin(j\pi u)}{j}]\mathbf{V}_k =$$

$$= \frac{\sqrt{2}}{\pi}\sum_{j=1}^{M}[\sum_{k=1}^{M}\mathbf{a}'_m(\delta_{jk}\mathbf{I}_{m^2}-\mathbf{P}_{jk})\mathbf{V}_k]\frac{\sin(j\pi u)}{j}+\frac{\sqrt{2}}{\pi}\sum_{k=1}^{M}[\sum_{j=M+1}^{\infty}\mathbf{a}'_m(\delta_{jk}\mathbf{I}_{m^2}-\mathbf{P}_{jk})\frac{\sin(j\pi u)}{j}]\mathbf{V}_k \ . \tag{3.23}$$

The first summand of the last line in expression (3.23) may be chosen as a candidate for $A^M(u)$. This can be rewritten as

$$A^M(u) = \boldsymbol{\alpha}'_M(u)(\mathbf{I}_M \otimes \mathbf{a}'_m)(\mathbf{I}_{Mm^2} - \mathbf{\Pi}_M)\begin{pmatrix}\mathbf{V}_1\\ \mathbf{V}_2\\ \vdots\\ \mathbf{V}_M\end{pmatrix} \ . \tag{3.24}$$

On the other hand, collecting the remaining terms together

$$R^M(u) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{M}[\sum_{j=M+1}^{\infty}\mathbf{a}'_m(\delta_{jk}\mathbf{I}_{m^2} - \mathbf{P}_{jk})\frac{\sin(j\pi u)}{j}]\mathbf{V}_k + \frac{\sqrt{2}}{\pi}\sum_{k=M+1}^{\infty}\mathbf{a}'_m\mathbf{p}_k(u)\mathbf{V}_k \ . \tag{3.25}$$

**Third part.** The rest of the proof is just a matter of verifying that the conditions **(C.1)**–**(C.2)**–**(C.3)** of Lemma 2.6.1 hold for $M \ge M_0 = p + q$.

**(C.1)** For $r \geq 1$ and $u_1$, $u_2$, …, $u_r$ in $[0,1]$, from (3.21) it can be written

$$\begin{pmatrix} A_n^d(u_1) \\ A_n^d(u_2) \\ \vdots \\ A_n^M(u_r) \end{pmatrix} = \mathbf{M}(\mathbf{I}_M \otimes \mathbf{a}_m')\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}(\mathbf{I}_{Mm^2} - \mathbf{\Pi}_M)\sqrt{n}\mathcal{W}^{-1/2}\mathbf{G}_M \ ,$$

where $\mathbf{M}$ is a $r \times M$ constant matrix whose $jth$ row is $\boldsymbol{\alpha}_M'(u_j)$, $j = 1, \ldots, r$. As explained in Chapter 2, $\widehat{\mathbf{\Sigma}} \xrightarrow{P} \mathbf{\Sigma} > \mathbf{0}$. Then, $\widehat{\mathbf{W}}^{-1/2} \xrightarrow{P} \mathcal{W}^{-1/2}$. Accordingly, from (3.24), Proposition 2.6.1, and Slutsky's theorem, as $n \to \infty$

$$\begin{pmatrix} A_n^M(u_1) \\ A_n^M(u_2) \\ \vdots \\ A_n^M(u_r) \end{pmatrix} \xrightarrow{D} \mathbf{M}(\mathbf{I}_M \otimes \mathbf{a}_m')(\mathbf{I}_{Mm^2} - \mathbf{\Pi}_M) \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix} \stackrel{D}{=} \begin{pmatrix} A^M(u_1) \\ A^M(u_2) \\ \vdots \\ A^M(u_r) \end{pmatrix} \ . \ \blacksquare$$

**(C.2)** Since $A_n^M(0) = 0$, condition *(i)* of Theorem 7.3 in Billingsley (1976, p. 82) holds. On the other hand, by (3.21) and the Cauchy-Schwarz inequality, $|A_n^M(u) - A_n^M(v)| \leq Z_n^M\,|u - v|$, where from part **(C.1)**

$$Z_n^M = \sqrt{2M}\,\|(\mathbf{I}_M \otimes \mathbf{a}_m')\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}(\mathbf{I}_{Mm^2} - \mathbf{\Pi}_M)\sqrt{n}\mathcal{W}^{-1/2}\mathbf{G}_M\| = O_P(1) \ .$$

Using similar arguments to those in **(C.2)** of Theorem 2.6.1, condition *(ii)* in Billingsley (1976, p. 82) also holds. Tightness of $\{A_n^M(u) : 0 \leq u \leq 1\}$ then follows. $\blacksquare$

**(C.3)** Using definition (3.14) in (3.25), after some algebra it can be written $R^M(u) = \sum_{i=1}^3 R_i^M(u)$, $0 \leq u \leq 1$, where

$$R_1^M(u) = \frac{\sqrt{2}}{\pi}\sum_{j=M+1}^{\infty} \boldsymbol{\beta}_{j,M}\frac{\sin(j\pi u)}{j} \ ; \tag{3.26}$$

$$R_2^M(u) = \frac{\sqrt{2}}{\pi}\sum_{j=1}^{M} \boldsymbol{\alpha}_{j,M}\frac{\sin(j\pi u)}{j} \ ; \tag{3.27}$$

$$R_3^M(u) = \frac{\sqrt{2}}{\pi}\sum_{k=M+1}^{\infty} \mathbf{a}_m'\mathbf{V}_k\frac{\sin(k\pi u)}{k} \ ; \tag{3.28}$$

$\boldsymbol{\beta}_{j,M} = -\sum_{k=1}^{\infty} \mathbf{a}_m'\mathbf{P}_{jk}\mathbf{V}_k$, $j \geq M+1$; and $\boldsymbol{\alpha}_{j,M} = -\sum_{k=M+1}^{\infty} \mathbf{a}_m'\mathbf{P}_{jk}\mathbf{V}_k$, $1 \leq j \leq M$. Thus, to get the condition $\limsup_M \Pr(\sup_{0\leq u\leq 1}|R^M(u)| > \varepsilon) = 0$, it suffices to check that $\limsup_M \Pr(\sup_{0\leq u\leq 1}|R_i^M(u)| > \varepsilon) = 0$, $i = 1, 2, 3$. Each of these limits is studied in Appendix 3.3.

On the other hand, we prove that

$$\limsup_{M} [\limsup_{n} \Pr(\sup_{0\le u\le 1} |R_n^M(u)| > \varepsilon)] = 0 \ , \tag{3.29}$$

by showing that each summand at the right-hand side of decomposition (4.34), $R_n^M(u) = P_n^M(u) + Q_n^M(u) + S_n^M(u)$, satisfies the corresponding limit condition. The terms $P_n^M(u)$ and $Q_n^M(u)$ are treated next. Derivations for $S_n^M(u)$ are more elaborated, and are given in Appendix 3.4.

By Cauchy-Schwarz inequality,

$$\sup_{0\le u\le 1} |P_n^M(u)| \le c\|(\mathbf{I}_M \otimes \mathbf{a}'_m)\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}(\mathbf{\Pi}_M - \mathbf{P}_M)\sqrt{n}\mathcal{W}^{-1/2}\mathbf{G}_M\| \ , \tag{3.30}$$

where $c = (\sqrt{2}/\pi)(\sum_{j=1}^{\infty} 1/j^2)^{1/2}$. Proceeding as in **(C.1)** it follows that, for each $M$,

$$(\mathbf{I}_M \otimes \mathbf{a}'_m)\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}(\mathbf{\Pi}_M - \mathbf{P}_M)\sqrt{n}\mathcal{W}^{-1/2}\mathbf{G}_M \xrightarrow{D} (\mathbf{I}_M \otimes \mathbf{a}'_m)(\mathbf{\Pi}_M - \mathbf{P}_M)\begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix}$$

as $n \to \infty$. From Appendix 3.1, $\|\mathbf{\Pi}_M - \mathbf{P}_M\| \to 0$ as $M \to \infty$. Define now

$$P^M = c\|(\mathbf{I}_M \otimes \mathbf{a}'_m)(\mathbf{\Pi}_M - \mathbf{P}_M)\begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix}\| \ . \tag{3.31}$$

From inequality (3.30), it is obtained that, for each $M$,

$$\limsup_{n} \Pr(\sup_{0\le u\le 1} |P_n^M(u)| > \varepsilon) \le$$

$$\le \limsup_{n} \Pr(c\|(\mathbf{I}_M \otimes \mathbf{a}'_m)\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}(\mathbf{\Pi}_M - \mathbf{P}_M)\sqrt{n}\mathcal{W}^{-1/2}\mathbf{G}_M\| > \varepsilon) =$$

$$= \Pr(P^M > \varepsilon) \le \mathrm{E}[(P^M)^2]/\varepsilon^2 \ .$$

Therefore, to finish the proof of condition **(C.3)** for $P_n^M(u)$, it is enough to establish that $\mathrm{E}[(P^M)^2] \to 0$ as $M \to \infty$. From its definition in (3.31), it is obtained that

$$\mathrm{E}[(P^M)^2] =$$

$$= c^2\mathrm{E}[(\mathbf{V}_1{}', \mathbf{V}_2{}', \ldots, \mathbf{V}_M{}')\,(\mathbf{\Pi}_M - \mathbf{P}_M)(\mathbf{I}_M \otimes \mathbf{a}_m)(\mathbf{I}_M \otimes \mathbf{a}'_m)(\mathbf{\Pi}_M - \mathbf{P}_M)\begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix}] =$$

$$= c^2 \mathrm{tr}[(\mathbf{\Pi}_M - \mathbf{P}_M)(\mathbf{I}_M \otimes \mathbf{a}_m \mathbf{a}_m')(\mathbf{\Pi}_M - \mathbf{P}_M)] \ ,$$

where $\mathbf{I}_M \otimes \mathbf{a}_m \mathbf{a}_m'$ is a $Mm^2 \times Mm^2$ projection matrix. Thus, it follows from the above that $\mathrm{E}[(P^M)^2] \leq c^2 \mathrm{tr}[(\mathbf{\Pi}_M - \mathbf{P}_M)(\mathbf{\Pi}_M - \mathbf{P}_M)] = c^2 \|\mathbf{\Pi}_M - \mathbf{P}_M\|^2 \to 0$, as $M \to \infty$.

By the orthogonality conditions of (2.28), for $M$ large enough $\mathbf{Z}_M' \mathcal{W}^{-1} \widehat{\mathbf{G}}_M = O_P(1/n)$. Thus, the term

$$Q_n^M(u) = \boldsymbol{\alpha}_M'(u)(\mathbf{I}_M \otimes \mathbf{a}_m')\widehat{\mathbf{W}}^{-1/2} \mathcal{W}^{1/2} \mathbf{P}_M \sqrt{n} \mathcal{W}^{-1/2} \widehat{\mathbf{G}}_M$$

in (4.34) is such that $\sup_{0 \leq u \leq 1} |Q_n^M(u)| = O_P(1/\sqrt{n}) = o_P(1)$. As a result, for large values of $M$, $\limsup_n \Pr(\sup_{0 \leq u \leq 1} |Q_n^M(u)| > \varepsilon) = 0$. Hence,

$$\limsup_M \, [\limsup_n \Pr( \sup_{0 \leq u \leq 1} |Q_n^M(u)| > \varepsilon)] = 0 \ . \ \blacksquare$$

## 3.5 Consequences

By Theorem 3.4.1 and the continuous mapping theorem, the asymptotic distribution of any continuous functional $H[\widehat{W}_n^m(u)]$ of the residual process of (1.28) is given by $H[G^m(u)]$, where $\{G^m(u) : 0 \leq u \leq 1\}$ is the Gaussian process of (3.15). As seen in section 3.2.2, the covariance function of $\{G^m(u) : 0 \leq u \leq 1\}$ depends on the unknown parameters $(\mathbf{\Phi}, \mathbf{\Theta}, \mathbf{\Sigma})$ of model (1.1). Therefore, assessing for goodness-of-fit purposes the significance of an observed value of $H[\widehat{W}_n^m(u)]$ with $H[G^m(u)]$ is not feasible, because the distribution of $H[G^m(u)]$ is parametric.

Since the residuals $\widehat{\boldsymbol{\varepsilon}}_t$ are estimators of the errors $\boldsymbol{\varepsilon}_t$, it is natural to expect that the behavior of the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (1.28) will be similar to that of its error counterpart $\{W_n^m(u) : 0 \leq u \leq 1\}$ of (2.44). By Theorem 2.6.2, $W_n^m(u) \to_\omega B(u)$. Therefore, a reasonable alternative could to be to assess the significance of $H[\widehat{W}_n^m(u)]$ with the quantiles of the pivotal distribution of $H[B(u)]$.

However, from expression (3.2) the covariance function $\gamma^m(u,v)$ of $\{G^m(u) : 0 \leq u \leq 1\}$ is smaller than the covariance function of the Brownian bridge in the Loewner sense. That is, for any function $l(.)$ such that the integral is well defined,

$$\int_0^1 \int_0^1 [\min(u,v) - uv - \gamma^m(u,v)] l(u) l(v) du dv =$$

$$= \frac{1}{2\pi^2 m} \int_0^1 \int_0^1 \mathbf{g}^m(\pi u)' \mathbf{I}^{-1}(\mathbf{\Lambda}) \mathbf{g}^m(\pi v) l(u) l(v) du dv =$$

$$= \frac{1}{2\pi^2 m} [\int_0^1 l(u) \mathbf{g}^m(\pi u) du]' \mathbf{I}^{-1}(\mathbf{\Lambda}) [\int_0^1 l(v) \mathbf{g}^m(\pi v) dv] \geq 0 \ .$$

Therefore, by an inequality due to Anderson (1955) it follows that

$$\Pr[G^m(u) \in S] \geq P[B(u) \in S] \ , \tag{3.32}$$

where $S$ is any convex and symmetric Borel set in $C[0,1]$.

As an application of (3.32), consider the Kolmogorov-Smirnov ($KS$), $H[f(u)] = \sup_{0 \leq u \leq 1} |f(u)|$, and Cramér-von Mises ($CVM$), $H[f(u)] = \int_0^1 f^2(u)du$, functionals, where $f = f(u) \in C[0,1]$. It is straightforward to check that the sets $\{f \in C[0,1] : \sup_{0 \leq u \leq 1} |f(u)| < c\}$ and $\{f \in C[0,1] : \int_0^1 f^2(u)du < c\}$, where $c > 0$, are convex and symmetric. Thus, as a consequence of inequality (3.32), it is obtained that for both the $KS$ and $CVM$ functionals,

$$\Pr\{H[G^m(u)] \geq H_\alpha[B(u)]\} \leq \Pr\{H[B(u)] \geq H_\alpha[B(u)]\} = \alpha \ , \tag{3.33}$$

where $H_\alpha[B(u)]$ is the $(1-\alpha)-$ quantile of the distribution of $H[B(u)]$. From (3.33), the rejection criterion

$$H[\widehat{W}_n^m(u)] \geq H_\alpha[B(u)] \ , \tag{3.34}$$

will tend to be conservative for the null hypothesis specified by model (1.1). This is because, for $n$ large,

$$\Pr\{H[\widehat{W}_n^m(u)] \geq H_\alpha[B(u)]\} \cong \Pr\{H[G^m(u)] \geq H_\alpha[B(u)]\} \leq \alpha \ . \tag{3.35}$$

This phenomenon, that can be very severe, will be investigated later in the simulation experiments of chapter 5.

As an alternative to replacing $H[G^m(u)]$ by $H[B(u)]$, that produces the distortion of (5.52), a goodness-of-fit process based on a transformed correlation matrix sequence is proposed in the next chapter, that converges to the Brownian bridge.

## Appendix 3.1: Properties of the matrices $\mathbf{P}_{jk}$

In what follows, the Euclidean norm $\|\mathbf{A}\|$ of a square matrix $\mathbf{a}_m = (a_{ij})$ will be considered. By definition, $\|\mathbf{A}\| = [\mathrm{tr}(\mathbf{A}\mathbf{A}')]^{1/2} = (\sum_i \sum_j a_{ij}^2)^{1/2}$. By the assumptions on the structure of the roots of the polynomial equations $|\mathbf{\Phi}(z)| = 0$ and $|\mathbf{\Theta}(z)| = 0$ made after the definition of model (1.1), and proceeding as in Theorem 11.3.1 in Brockwell and Davis (1991, p. 408), it follows that

$$\max\{\|\mathbf{\Omega}_j\|, \|\mathbf{L}_j\|\} \leq ab^j \ , \ j \geq 0 \ , \tag{3.36}$$

where $a > 0$; $0 < b < 1$; and $\{\mathbf{\Omega}_j : j \geq 0\}$ and $\{\mathbf{L}_j : j \geq 0\}$ are the $m \times m$ coefficients of the series expansions $\mathbf{\Phi}^{-1}(z)\mathbf{\Theta}(z) = \sum_{j=0}^{\infty} \mathbf{\Omega}_j z^j$ and $\mathbf{\Theta}^{-1}(z) = \sum_{j=0}^{\infty} \mathbf{L}_j z^j$, respectively. Put also $\mathbf{G}_k = \sum_{j=0}^{k} (\mathbf{\Sigma}\mathbf{\Omega}_j' \otimes \mathbf{L}_{k-j})$ and $\mathbf{F}_k = \mathbf{\Sigma} \otimes \mathbf{L}_k$, $k \geq 0$.

Using the bound $\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$, it can be obtained

$$\|(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\mathbf{G}_k\| \leq \sum_{j=0}^{k} \|\mathbf{\Sigma}^{1/2}\mathbf{\Omega}_j' \otimes \mathbf{\Sigma}^{-1/2}\mathbf{L}_{k-j}\| =$$

$$= \sum_{j=0}^{k} \|(\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{-1/2})(\mathbf{\Omega}_j' \otimes \mathbf{L}_{k-j})\| \leq \|(\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{-1/2})\| \sum_{j=0}^{k} \|\mathbf{\Omega}_j' \otimes \mathbf{L}_{k-j}\| \ . \tag{3.37}$$

On the other hand, using the formula $\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})$, it can be checked that

$$\|\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{-1/2}\|^2 =$$

$$= \mathrm{tr}[(\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{-1/2})(\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{-1/2})] = \mathrm{tr}(\mathbf{\Sigma} \otimes \mathbf{\Sigma}^{-1}) = \mathrm{tr}(\mathbf{\Sigma})\mathrm{tr}(\mathbf{\Sigma}^{-1}) \ . \tag{3.38}$$

Moreover, using inequality (3.36),

$$\|\mathbf{\Omega}_j' \otimes \mathbf{L}_{k-j}\|^2 = \mathrm{tr}[(\mathbf{\Omega}_j' \otimes \mathbf{L}_{k-j})(\mathbf{\Omega}_j \otimes \mathbf{L}_{k-j}')] = \mathrm{tr}[(\mathbf{\Omega}_j'\mathbf{\Omega}_j) \otimes (\mathbf{L}_{k-j}\mathbf{L}_{k-j}')] =$$

$$= \mathrm{tr}(\mathbf{\Omega}_j'\mathbf{\Omega}_j)\mathrm{tr}(\mathbf{L}_{k-j}\mathbf{L}_{k-j}') = \|\mathbf{\Omega}_j\|^2\|\mathbf{L}_{k-j}\|^2 \leq a^4 b^{2j} b^{2(k-j)} = a^4 b^{2k}. \tag{3.39}$$

Combining expressions (3.36) – (3.37) – (3.38) – (3.39) leads finally, after some algebra, to the inequalities

$$\max\{\|(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\mathbf{G}_k\|, \|(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\mathbf{F}_k\|\} \leq cd^k \ , \ k \geq 0 \ , \tag{3.40}$$

for some constants $c > 0$, and $0 < d < 1$.

From (3.6), $\mathbf{P}_{jk} = \mathbf{\Xi}_j' \mathbf{I}^{-1}(\mathbf{\Lambda})\mathbf{\Xi}_k$, $j, k \geq 1$, where using the notation of (2.23),

$$\mathbf{\Xi}_k' = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})(\mathbf{G}_{k-1}, \ldots, \mathbf{G}_{k-p}\, ; \mathbf{F}_{k-1}, \ldots, \mathbf{F}_{k-q}) \ ,$$

is the $k$th $m^2 \times m^2(p+q)$ row block of the matrix $\mathcal{W}^{-1/2}\mathbf{Z}_M$, $k = 1, \ldots, M$. Therefore, from inequality (3.40) it follows after some algebra that

$$\|\mathbf{P}_{jk}\| \leq f g^{j+k} \;, \qquad j, k \geq 1 \;, \tag{3.41}$$

where $f > 0$, and $0 < g < 1$. Proceeding as in (3.41), it can be also obtained that the blocks of the matrix $\mathbf{P}_M = (\mathbf{P}_{jk,M} \colon 1 \leq j, k \leq M)$ satisfy

$$\|\mathbf{P}_{jk,M}\| \leq f_M g^{j+k} \;, \qquad 1 \leq j, k \leq M \;, \tag{3.42}$$

where $0 < f_M \to f$ as $M \to \infty$; and

$$\|\mathbf{P}_{jk} - \mathbf{P}_{jk,M}\| \leq h_M g^{j+k} \;, \qquad 1 \leq j, k \leq M \;, \tag{3.43}$$

where $0 < h_M \to 0$ as $M \to \infty$.

Some consequences of the inequalities (3.41) – (3.42) – (3.43) for the behavior of the blocks of the $Mm^2 \times Mm^2$ matrices $\mathbf{\Pi}_M = (\mathbf{P}_{jk} \colon 1 \leq j, k \leq M)$ and $\mathbf{P}_M = (\mathbf{P}_{jk,M} \colon 1 \leq j, k \leq M)$ are listed below:

**(a)** $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \|\mathbf{P}_{jk}\| \leq f \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} g^{j+k} = f \sum_{j=1}^{\infty} g^j \sum_{k=1}^{\infty} g^k < \infty$.

**(b)** $|\mathbf{a}_m' \mathbf{P}_{jk} \mathbf{a}_m| \leq \|\mathbf{a}_m'\| \|\mathbf{P}_{jk}\| \|\mathbf{a}_m\| \leq f g^{j+k}$, $j, k \geq 1$.

**(c)** $\sup_M \sum_{j=1}^{M} \sum_{k=1}^{M} \|\mathbf{P}_{jk,M}\| \leq (\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} g^{j+k}) \sup_M f_M < \infty$.

**(d)** $\sup_{1 \leq j,k \leq M} \|\mathbf{P}_{jk} - \mathbf{P}_{jk,M}\| \leq h_M \to 0$.

**(e)** $\|\mathbf{\Pi}_M - \mathbf{P}_M\|^2 = \sum_{j=1}^{M} \sum_{k=1}^{M} \|\mathbf{P}_{jk} - \mathbf{P}_{jk,M}\|^2 \leq h_M^2 (\sum_{j=1}^{\infty} g^{2j})^2 \to 0$, as $M \to \infty$.

By writing finally

$$\mathbf{P}_{jk,M} = \sum_{s=1}^{M} \mathbf{P}_{js,M} \mathbf{P}_{sk,M} = \sum_{s=1}^{M} (\mathbf{P}_{js,M} - \mathbf{P}_{js})(\mathbf{P}_{sk,M} - \mathbf{P}_{sk})$$

$$+ \sum_{s=1}^{M} \mathbf{P}_{js} (\mathbf{P}_{sk,M} - \mathbf{P}_{sk}) + \sum_{s=1}^{M} (\mathbf{P}_{js,M} - \mathbf{P}_{js}) \mathbf{P}_{sk} + \sum_{s=1}^{M} \mathbf{P}_{js} \mathbf{P}_{sk} \;,$$

the results above can be used to justify the limit expression (3.7). ∎

## Appendix 3.2: Derivation of identity (3.12)

The $m \times m$ spectral density matrix (3.4) of the process of (1.1) can be written

$$\mathbf{f}(\omega, \mathbf{\Lambda}) = \frac{1}{2\pi}\,\mathbf{k}(\omega, \mathbf{\Lambda})\mathbf{\Sigma}\mathbf{k}^*(\omega, \mathbf{\Lambda}) \;, \quad -\pi \le \omega \le \pi \;, \tag{3.44}$$

where $\mathbf{k}(\omega, \mathbf{\Lambda}) = \mathbf{\Phi}^{-1}(e^{i\omega})\mathbf{\Theta}(e^{i\omega})$ is as considered in section 2.3; $\mathbf{k}^*(\omega, \mathbf{\Lambda})$ is the conjugate transpose; and $\mathbf{\Lambda} = \text{vec}(\mathbf{\Phi}, \mathbf{\Theta}) = [\lambda_1, \ldots, \lambda_{m^2(p+q)}]'$. Taking partial derivatives in expression (3.44), it follows that

$$\frac{\partial\, \mathbf{f}(\omega, \mathbf{\Lambda})}{\partial \lambda_i} = \frac{1}{2\pi}\,[\frac{\partial \mathbf{k}(\omega, \mathbf{\Lambda})}{\partial \lambda_i}\mathbf{\Sigma}\mathbf{k}^*(\omega, \mathbf{\Lambda}) + \mathbf{k}(\omega, \mathbf{\Lambda})\mathbf{\Sigma}\frac{\partial \mathbf{k}^*(\omega, \mathbf{\Lambda})}{\partial \lambda_i}] \;.$$

Consequently, it can be written

$$\mathbf{f}^{-1}(\omega, \mathbf{\Lambda})\,\frac{\partial\, \mathbf{f}(\omega, \mathbf{\Lambda})}{\partial \lambda_i} =$$

$$= \mathbf{k}^{*^{-1}}(\omega, \mathbf{\Lambda})\mathbf{\Sigma}^{-1}\mathbf{k}^{-1}(\omega, \mathbf{\Lambda})\,\frac{\partial \mathbf{k}(\omega, \mathbf{\Lambda})}{\partial \lambda_i}\,\mathbf{\Sigma}\,\mathbf{k}^*(\omega, \mathbf{\Lambda}) + \mathbf{k}^{*^{-1}}(\omega, \mathbf{\Lambda})\,\frac{\partial \mathbf{k}^*(\omega, \mathbf{\Lambda})}{\partial \lambda_i} \;. \tag{3.45}$$

Using (3.45), it is obtained that

$$\text{tr}[\mathbf{f}^{-1}(\omega, \mathbf{\Lambda})\mathbf{f}(\omega, \mathbf{\Lambda})/\partial \lambda_i] = \text{tr}\,[\mathbf{A}(\omega, \lambda_i) + \mathbf{A}^*(\omega, \lambda_i)] \;, \tag{3.46}$$

where $\mathbf{A}(\omega, \lambda_i) = \mathbf{k}^{-1}(\omega, \mathbf{\Lambda})\,[\partial \mathbf{k}(\omega, \mathbf{\Lambda})/\partial \lambda_i]$, $i = 1, \ldots, m^2(p+q)$.

Consider now the $m \times m$ coefficients $\{\mathbf{\Omega}_k : k \ge 0\}$ and $\{\mathbf{L}_k : k \ge 0\}$ of the series expansions $\mathbf{\Phi}^{-1}(z)\mathbf{\Theta}(z) = \sum_{k=0}^{\infty} \mathbf{\Omega}_k z^k$ and $\mathbf{\Theta}^{-1}(z) = \sum_{k=0}^{\infty} \mathbf{L}_k z^k$, respectively. Put also $\mathbf{G}_j = \sum_{k=0}^{j}(\mathbf{\Sigma}\mathbf{\Omega}_k' \otimes \mathbf{L}_{j-k})$, and $\mathbf{F}_j = \mathbf{\Sigma} \otimes \mathbf{L}_j$, $j \ge 0$. Using the definition of $\mathbf{a}_m = \text{vec}(\mathbf{I}_m)/\sqrt{m}$ given in (2.33), write

$$\mathbf{g}_j' = \mathbf{a}_m'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\mathbf{G}_j = \mathbf{a}_m'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\sum_{k=0}^{j}(\mathbf{\Sigma}\mathbf{\Omega}_k' \otimes \mathbf{L}_{j-k}) =$$

$$= \mathbf{a}_m'\sum_{k=0}^{j}(\mathbf{\Sigma}^{1/2}\mathbf{\Omega}_k' \otimes \mathbf{\Sigma}^{-1/2}\mathbf{L}_{j-k}) = \frac{1}{\sqrt{m}}\sum_{k=0}^{j}[\text{vec}(\mathbf{L}_{j-k}'\mathbf{\Omega}_k')]' \;, \quad j \ge 0 \;; \tag{3.47}$$

and

$$\mathbf{f}_j' = \mathbf{a}_m'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\mathbf{F}_j = \mathbf{a}_m'(\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{-1/2}\mathbf{L}_j) = [\text{vec}(\mathbf{L}_j')]'/\sqrt{m} \;, \quad j \ge 0 \;. \tag{3.48}$$

Taking into account the $m^2 \times m^2(p+q)$ row blocks defined in (2.23),

$$\mathbf{\Xi}_j' = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})(\mathbf{G}_{j-1}, \ldots, \mathbf{G}_{j-p}\,; \mathbf{F}_{j-1}, \ldots, \mathbf{F}_{j-q}) \;,$$

define the $1 \times m^2(p+q)$ vector

$$\boldsymbol{\eta}_j' = \mathbf{a}_m'\boldsymbol{\Xi}_j' = \mathbf{a}_m'(\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2})(\mathbf{G}_{j-1}, \ldots, \mathbf{G}_{j-p}; \mathbf{F}_{j-1}, \ldots, \mathbf{F}_{j-q}) =$$

$$= (\mathbf{g}_{j-1}', \ldots, \mathbf{g}_{j-p}'; \mathbf{f}_{j-1}', \ldots, \mathbf{f}_{j-q}') \ , \quad j \geq 0 \ , \tag{3.49}$$

whose $1 \times m^2$ components are obtained from expressions (3.47)−(3.48) above.

Using now expression (2.66) in Appendix 2.1,

$$\mathbf{A}(\omega, \phi_{jk,r}) = \sum_{u=0}^{\infty}\sum_{v=0}^{\infty} \mathbf{L}_u \mathbf{e}_j \mathbf{e}_k' \boldsymbol{\Omega}_v e^{i(r+u+v)\omega} \ ,$$

and the identities $\mathrm{tr}(\mathbf{L}_u\mathbf{e}_j\mathbf{e}_k'\boldsymbol{\Omega}_v) = \mathrm{tr}(\mathbf{e}_k'\boldsymbol{\Omega}_v\mathbf{L}_u\mathbf{e}_j) = \mathrm{tr}(\mathbf{e}_j'\mathbf{L}_u'\boldsymbol{\Omega}_v'\mathbf{e}_k)$, it is obtained

$$\int_0^{\pi u} \mathrm{tr}[\mathbf{A}(\omega, \phi_{jk,r}) + \mathbf{A}^*(\omega, \phi_{jk,r})]d\omega =$$

$$2\sum_{u=0}^{\infty}\sum_{v=0}^{\infty} \mathrm{tr}(\mathbf{e}_j'\mathbf{L}_u'\boldsymbol{\Omega}_v'\mathbf{e}_k) \int_0^{\pi u} \cos[(r+u+v)\omega]d\omega \ . \tag{3.50}$$

Introducing the index $K = r + u + v$ in (3.50), it follows that

$$\int_0^{\pi u} \mathrm{tr}[\mathbf{A}(\omega, \phi_{jk,r}) + \mathbf{A}^*(\omega, \phi_{jk,r})]d\omega = 2\sum_{K=1}^{\infty}[\sum_{v=0}^{K-r} \mathrm{tr}(\mathbf{e}_j'\mathbf{L}_{K-r-v}'\boldsymbol{\Omega}_v'\mathbf{e}_k)]\frac{\sin(K\pi u)}{K} \ , \tag{3.51}$$

where from (3.47)

$$\sum_{v=0}^{K-r} \mathrm{tr}(\mathbf{e}_j'\mathbf{L}_{K-r-v}'\boldsymbol{\Omega}_v'\mathbf{e}_k) = \sum_{v=0}^{K-r} \mathrm{tr}(\mathbf{e}_k\mathbf{e}_j'\mathbf{L}_{K-r-v}'\boldsymbol{\Omega}_v') =$$

$$= [\mathrm{vec}(\mathbf{e}_j\mathbf{e}_k')]' \sum_{v=0}^{K-r} \mathrm{vec}[\mathbf{L}_{K-r-v}'\boldsymbol{\Omega}_v'] = \sqrt{m}\,[\mathrm{vec}(\mathbf{e}_j\mathbf{e}_k')]'\mathbf{g}_{K-r} \ . \tag{3.52}$$

On the other hand, after some algebra it can be written

$$\mathbf{A}(\omega, \theta_{jk,s}) = \sum_{u=0}^{\infty} \mathbf{L}_u\mathbf{e}_j\mathbf{e}_k' e^{i(s+u)\omega} \ .$$

Thus, taking into account that $\mathrm{tr}(\mathbf{L}_u\mathbf{e}_j\mathbf{e}_k') = \mathrm{tr}(\mathbf{e}_k\mathbf{e}_j'\mathbf{L}_u')$, it is obtained

$$\int_0^{\pi u} \mathrm{tr}[\mathbf{A}(\omega, \theta_{jk,s}) + \mathbf{A}^*(\omega, \theta_{jk,s})]d\omega = 2\sum_{u=0}^{\infty} \mathrm{tr}(\mathbf{e}_k\mathbf{e}_j'\mathbf{L}_u') \int_0^{\pi u} \cos[(s+u)\omega]d\omega \ . \tag{3.53}$$

Introducing now the index $K = s + u$, the identity below holds

$$\int_0^{\pi u} \mathrm{tr}[\mathbf{A}(\omega, \theta_{jk,s}) + \mathbf{A}^*(\omega, \theta_{jk,s})] d\omega = 2 \sum_{K=1}^{\infty} \mathrm{tr}(\mathbf{e}_k \mathbf{e}_j' \mathbf{L}_{K-s}') \frac{\sin(K\pi u)}{K} \ , \tag{3.54}$$

where, from (3.48), $\mathrm{tr}(\mathbf{e}_k \mathbf{e}_j' \mathbf{L}_{K-s}') = [\mathrm{vec}(\mathbf{e}_j \mathbf{e}_k')]' \mathrm{vec}(\mathbf{L}_{K-s}') = \sqrt{m}\, \mathrm{vec}(\mathbf{e}_j \mathbf{e}_k')]' \mathbf{f}_{K-s}$.

As a consequence of both (3.50)−(3.51)−(3.52) and (3.53)−(3.54), the $m^2(p+q)\times 1$ function $\mathbf{g}^m(\pi u)$ of (3.3) is such that

$$\frac{\mathbf{g}^m(\pi u)}{2\sqrt{m}} = \sum_{j=1}^{\infty} \frac{\sin(j\pi u)}{j} \boldsymbol{\eta}_j \ , \quad 0 \le u \le 1 \ , \tag{3.55}$$

where $\boldsymbol{\eta}_j' = \mathbf{a}_m' \boldsymbol{\Xi}_j'$ is as defined in (3.49). From (3.55) and (3.6) it follows finally

$$\frac{1}{4m} \mathbf{g}^m(\pi u)' \mathbf{I}^{-1}(\boldsymbol{\Lambda}) \mathbf{g}^m(\pi v) = [\sum_{j=1}^{\infty} \frac{\sin(j\pi u)}{j} \ \boldsymbol{\eta}_j]' \, \mathbf{I}^{-1}(\boldsymbol{\Lambda}) \, [\sum_{k=1}^{\infty} \frac{\sin(k\pi v)}{k} \ \boldsymbol{\eta}_k] =$$

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} [\boldsymbol{\eta}_j' \, \mathbf{I}^{-1}(\boldsymbol{\Lambda}) \, \boldsymbol{\eta}_k] \, \frac{\sin(j\pi u)}{j} \frac{\sin(k\pi v)}{k} \ , \quad 0 \le u, v \le 1 \ ,$$

where $\boldsymbol{\eta}_j' \, \mathbf{I}^{-1}(\boldsymbol{\Lambda}) \, \boldsymbol{\eta}_k = \mathbf{a}_m' \boldsymbol{\Xi}_j' \, \mathbf{I}^{-1}(\boldsymbol{\Lambda}) \, \boldsymbol{\Xi}_k \mathbf{a}_m = \mathbf{a}_m' \mathbf{P}_{jk} \mathbf{a}_m$, $j, k \ge 1$. This is the end of the justification of identity (3.12).

Notice finally that the components of $\boldsymbol{\eta}_j$ in (3.49) do not depend on $\boldsymbol{\Sigma}$. Accordingly, the coordinates $\mathbf{a}_m' \mathbf{P}_{jk} \mathbf{a}_m$, $j, k \ge 1$, and thus the covariance function $\gamma^m(u, v)$ of (3.11), depend on $\boldsymbol{\Sigma}$ only through the information matrix $\mathbf{I}(\boldsymbol{\Lambda})$. ■

## Appendix 3.3: Limits in the first part of condition (C.3) of Theorem 3.4.1

Recall the notation

$$\begin{aligned}\boldsymbol{\beta}_{j,M} &= -\sum_{k=1}^{\infty} \mathbf{a}_m' \mathbf{P}_{jk} \mathbf{V}_k \ , \quad j \geq M+1 \ ; \\ \boldsymbol{\alpha}_{j,M} &= -\sum_{k=M+1}^{\infty} \mathbf{a}_m' \mathbf{P}_{jk} \mathbf{V}_k \ , \quad 1 \leq j \leq M \ ;\end{aligned}$$

where $\{\mathbf{V}_k : k \geq 1\}$ are i.i.d. $N_{m^2}(\mathbf{0}, \mathbf{I}_{m^2})$ random vectors. It can be obtained that

$$\mathrm{E}(\boldsymbol{\beta}_{j,M}^2) = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \mathbf{a}_m' \mathbf{P}_{jk} \mathrm{E}(\mathbf{V}_k \mathbf{V}_l{}') \mathbf{P}_{lj} \mathbf{a}_m = \sum_{k=1}^{\infty} \mathbf{a}_m' \mathbf{P}_{jk} \mathbf{P}_{kj} \mathbf{a}_m = \mathbf{a}_m' \mathbf{P}_{jj} \mathbf{a}_m \ , \quad j \geq M+1 \ .$$

From Appendix 3.1, $|\mathbf{a}_m' \mathbf{P}_{jj} \mathbf{a}_m| \leq \|\mathbf{a}_m'\| \|\mathbf{P}_{jj}\| \|\mathbf{a}_m\| \leq f g^{2j}$, $j \geq 1$, where $f > 0$, and $0 < g < 1$. Therefore, applying Cauchy-Schwarz inequality in definition (3.26),

$$R_1^M = \sup_{0 \leq u \leq 1} |R_1^M(u)| \leq \frac{\sqrt{2}}{\pi} \Big(\sum_{j=M+1}^{\infty} 1/j^2\Big)^{1/2} \Big(\sum_{j=M+1}^{\infty} \boldsymbol{\beta}_{j,M}^2\Big)^{1/2} \ ,$$

where $\mathrm{E}[(R_1^M)^2] \leq (2f/\pi^2)(\sum_{j=M+1}^{\infty} 1/j^2)(\sum_{j=M+1}^{\infty} g^{2j}) \to 0$ as $M \to \infty$. Hence,

$$\limsup_M \Pr\Big(\sup_{0 \leq u \leq 1} |R_1^M(u)| > \varepsilon\Big) = \limsup_M \Pr(R_1^M > \varepsilon) \leq \limsup_M \mathrm{E}[(R_1^M)^2]/\varepsilon^2 = 0 \ .$$

On the other hand,

$$\mathrm{E}(\boldsymbol{\alpha}_{j,M}^2) = \sum_{k=M+1}^{\infty}\sum_{l=M+1}^{\infty} \mathbf{a}_m' \mathbf{P}_{jk} \mathrm{E}(\mathbf{V}_k \mathbf{V}_l{}') \mathbf{P}_{lj} \mathbf{a}_m = \sum_{k=M+1}^{\infty} \mathbf{a}_m' \mathbf{P}_{jk} \mathbf{P}_{kj} \mathbf{a}_m \ , \quad 1 \leq j \leq M \ .$$

From inequality (3.41) in Appendix 3.1, it can be obtained $|\mathbf{a}_m' \mathbf{P}_{jk} \mathbf{P}_{kj} \mathbf{a}_m| \leq \|\mathbf{a}_m'\| \|\mathbf{P}_{jk}\| \|\mathbf{P}_{kj}\| \|\mathbf{a}_m\| \leq f^2 g^{2(j+k)}$, $k \geq M+1$, $1 \leq j \leq M$. Accordingly,

$$\mathrm{E}(\boldsymbol{\alpha}_{j,M}^2) \leq f^2 g^{2j} \sum_{k=M+1}^{\infty} g^{2k} = \frac{(fg)^2}{1-g^2}\, g^{2(j+M)} \ , \quad 1 \leq j \leq M \ .$$

If $R_2^M = \sup_{0 \leq u \leq 1} |R_2^M(u)|$, using Cauchy-Schwarz inequality in (3.27),

$$R_2^M = \sup_{0 \leq u \leq 1} |R_2^M(u)| \leq \frac{\sqrt{2}}{\pi} \Big(\sum_{j=1}^{M} 1/j^2\Big)^{1/2} \Big(\sum_{j=1}^{M} \boldsymbol{\alpha}_{j,M}^2\Big)^{1/2} \ ,$$

where

$$\mathrm{E}[(R_2^M)^2] \leq \frac{2(fg)^2}{\pi^2(1-g^2)} \Big[\Big(\sum_{j=1}^{M} 1/j^2\Big)\Big(\sum_{j=1}^{M} g^{2j}\Big)\Big]\, g^{2M} \to 0 \ ,$$

as $M \to \infty$. Proceeding as above, $\limsup_M \Pr(\sup_{0 \le u \le 1} |R_2^M(u)| > \varepsilon) = 0$.

Finally, since $\mathbf{a}_m$ is a unit $m^2 \times 1$ vector, $\{v_k = \mathbf{a}'_m \mathbf{V}_k : k \ge 1\}$ is a collection of i.i.d. $N(0,1)$ random variables. Therefore, $R_3^M(u)$ in (3.28) has the same structure that the term $R^M(u)$ that was studied in the first part of condition **(C.3)** of Theorem 2.6.1, where it was established that $\limsup_M \Pr(\sup_{0 \le u \le 1} |R^M(u)| > \varepsilon) = 0$. ∎

## Appendix 3.4: Limits in the second part of condition (C.3) of Theorem 3.4.1

From expression (2.38) in section 2.5,

$$\frac{1}{\sqrt{m}}\,\mathrm{tr}(\widehat{\mathbf{R}}_k) = \frac{1}{\sqrt{m}}\,\mathrm{tr}(\mathbf{R}_k) - \boldsymbol{\eta}_k'(\mathbf{Z}_M'\mathcal{W}^{-1}\mathbf{Z}_M)^{-1}\mathbf{U}_k + O_P(\frac{1}{n}) \ , \tag{3.56}$$

$k \geq 1$, where the $\boldsymbol{\eta}_k$ are as in definition (3.49) of Appendix 3.2, and $\mathbf{U}_k = (\mathbf{u}_{1,k}', \ldots, \mathbf{u}_{p,k}'; \mathbf{u}_{p+1,k}', \ldots, \mathbf{u}_{p+q,k}')'$ is a $m^2(p+q) \times 1$ vector with components

$$\mathbf{u}_{i,k} = \sum_{j=i}^{k} \mathbf{G}_{j-i}'(\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2})(\mathbf{C}_0^{-1/2} \otimes \mathbf{C}_0^{-1/2})\mathrm{vec}(\mathbf{C}_j') \ , \quad i = 1, \ldots, p, \tag{3.57}$$

where $\mathbf{G}_r'(\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2}) = \sum_{s=0}^{r}(\boldsymbol{\Omega}_s\boldsymbol{\Sigma} \otimes \mathbf{L}_{r-s}')(\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2}) = \sum_{s=0}^{r}(\boldsymbol{\Omega}_s\boldsymbol{\Sigma}^{1/2} \otimes \mathbf{L}_{r-s}'\boldsymbol{\Sigma}^{-1/2})$; and

$$\mathbf{u}_{p+I,k} = \sum_{j=I}^{k} \mathbf{F}_{j-I}'(\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2})(\mathbf{C}_0^{-1/2} \otimes \mathbf{C}_0^{-1/2})\mathrm{vec}(\mathbf{C}_j') \ , \quad I = 1, \ldots, p, \tag{3.58}$$

where $\mathbf{F}_r'(\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2}) = (\boldsymbol{\Sigma} \otimes \mathbf{L}_r')(\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2}) = \boldsymbol{\Sigma}^{1/2} \otimes \mathbf{L}_r'\boldsymbol{\Sigma}^{-1/2}$.

From (3.56), it can be written

$$S_n^M(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=M+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k} =$$

$$= C_n^M(u) - D_n^M(u) + F_n^M(u) \ , \quad 0 \leq u \leq 1 \ , \tag{3.59}$$

where

$$C_n^M(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=M+1}^{n-(P+1)} \frac{\mathrm{tr}(\mathbf{R}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k} \ ;$$

$$D_n^M(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=M+1}^{n-(P+1)} \boldsymbol{\eta}_k'(\mathbf{Z}_k'\mathcal{W}^{-1}\mathbf{Z}_k)^{-1}\mathbf{U}_k\frac{\sin(k\pi u)}{k} \ ;$$

and $F_n^d(u)$ is a remainder term suitable bounded in probability.

Condition **(C.3)** follows for $C_n^M(u)$ from the proofs of Theorems 2.6.1 and 2.6.2. For dealing with $D_n^M(u)$, define first $\omega_k = \boldsymbol{\eta}_k'(\mathbf{Z}_k'\mathcal{W}^{-1}\mathbf{Z}_k)^{-1}\mathbf{U}_k$ and their counterparts $\overline{\omega}_k = \boldsymbol{\eta}_k'(\mathbf{Z}_k'\mathcal{W}^{-1}\mathbf{Z}_k)^{-1}\overline{\mathbf{U}}_k$, $m^2(p+q) \leq k \leq n-(P+1)$, where the $\overline{\mathbf{U}}_k$ are similar to the $\mathbf{U}_k$, but with $\mathbf{C}_0 = \sum_{t=0}^{n}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'/n$ replaced by $\boldsymbol{\Sigma}$. Specifically, $\overline{\mathbf{U}}_k = (\overline{\mathbf{u}}_{1,k}', \ldots, \overline{\mathbf{u}}_{p,k}'; \overline{\mathbf{u}}_{p+1,k}', \ldots, \overline{\mathbf{u}}_{p+q,k}')'$ is a $m^2(p+q) \times 1$ vector such that $\overline{\mathbf{u}}_{i,k} =$

$\sum_{j=i}^{k} \mathbf{G}'_{j-i}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\text{vec}(\mathbf{C}'_j)$, $i = 1, \ldots, p$; and $\overline{\mathbf{u}}_{p+I,k} = \sum_{j=I}^{k} \mathbf{F}'_{j-i}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})$ $\text{vec}(\mathbf{C}'_j)$, $I = 1, \ldots, p$. Define

$$\overline{D}_n^M(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=M+1}^{n-(P+1)} \overline{\omega}_k \frac{\sin(k\pi u)}{k} \ . \quad 0 \le u \le 1 \ . \tag{3.60}$$

To check that $\limsup_M[\limsup_n \Pr(\sup_{0\le u\le 1} |D_n^M(u)| > \varepsilon)] = 0$, it is enough to proof the same condition for both $\overline{D}_n^M(u)$, and the difference $D_n^M(u) - \overline{D}_n^M(u)$.

To apply the results in Grenander and Rosenblatt (1957, Theorem 1, p. 188) for $\overline{D}_n^M(u)$ in (3.60), it is necessary to obtain a proper bound for

$$\text{E}(\overline{\omega}_k^2) = \boldsymbol{\eta}'_k(\mathbf{Z}'_k\mathcal{W}^{-1}\mathbf{Z}_k)^{-1}\text{E}(\overline{\mathbf{U}}_k\overline{\mathbf{U}}'_k)(\mathbf{Z}'_k\mathcal{W}^{-1}\mathbf{Z}_k)^{-1}\boldsymbol{\eta}_k \ , \tag{3.61}$$

where the matrix $\text{E}(\overline{\mathbf{U}}_k\overline{\mathbf{U}}'_k)$ is formed by blocks of the form $\text{E}[\overline{\mathbf{u}}_{H,k}\overline{\mathbf{u}}'_{L,k}]$, $H, L = 1, \ldots, p+q$. When $1 \le H, L \le p$,

$$\text{E}(\overline{\mathbf{u}}_{H,k}\overline{\mathbf{u}}'_{L,k}) =$$

$$= \sum_{J=H}^{k}\sum_{K=L}^{k} \mathbf{G}'_{J-H}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\text{E}\{\text{vec}(\mathbf{C}'_J)[\text{vec}(\mathbf{C}'_K)]'\}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{G}_{K-L} \ . \tag{3.62}$$

Proceeding as in Appendix 2.2,

$$E\{\text{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+J})[\text{vec}(\boldsymbol{\varepsilon}_u\boldsymbol{\varepsilon}'_{u+K})]'\} = \text{Cov}(\boldsymbol{\varepsilon}_{t+J}, \boldsymbol{\varepsilon}_{u+K}) \otimes E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_u) \ .$$

Therefore, from the independence of the sequence of random vectors $\{\boldsymbol{\varepsilon}_t\}$, and the identity $\text{vec}(\mathbf{C}'_k) = \mathbf{K}_{mm}\text{vec}(\mathbf{C}_k)$, where $\mathbf{K}_{mm}$ is the commutation matrix of order $m$ (Lütkepohl, 2005, Sec. A.12.2):

$$\text{E}\{\text{vec}(\mathbf{C}'_J)[\text{vec}(\mathbf{C}'_K)]'\} = \mathbf{K}_{mm}\text{E}\{\text{vec}(\mathbf{C}_J)[\text{vec}(\mathbf{C}_K)]'\}\mathbf{K}_{mm} =$$

$$= \frac{1}{n^2}\sum_{t=1}^{n-J}\sum_{u=1}^{n-K} \mathbf{K}_{mm}\text{E}\{\text{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+J})[\text{vec}(\boldsymbol{\varepsilon}_u\boldsymbol{\varepsilon}'_{u+K})]'\}\mathbf{K}_{mm} =$$

$$= \frac{1}{n^2}\sum_{t=1}^{n-J}\sum_{u=1}^{n-K} \mathbf{K}_{mm}\text{Cov}(\boldsymbol{\varepsilon}_{t+J}, \boldsymbol{\varepsilon}_{u+K}) \otimes E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_u)\mathbf{K}_{mm} =$$

$$= \frac{n-J}{n^2}\,\mathbf{K}_{mm}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{K}_{mm} = \frac{n-J}{n^2}\,(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \ , \quad J = K \ ;$$

and $\text{E}\{\text{vec}(\mathbf{C}'_J)[\text{vec}(\mathbf{C}'_K)]'\} = 0$ when $J \ne K$. Going back to (3.62),

$$\text{E}(\overline{\mathbf{u}}_{H,k}\overline{\mathbf{u}}'_{L,k}) = \sum_{J=\max(H,L)}^{k} \frac{n-J}{n^2}\,\mathbf{G}'_{J-H}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{G}_{J-L} \ .$$

The other blocks of $\mathrm{E}(\overline{\mathbf{U}}_k \overline{\mathbf{U}}'_k)$ can be treated similarly.

Putting all these things together, and using expression (3.36) of Appendix 3.1, it can be obtained that

$$\mathrm{E}(\overline{\varpi}_k^2) \leq \frac{a}{n}\, kb^k \ , \quad m^2(p+q) \leq k \ , \tag{3.63}$$

where $a > 0$, and $0 < b < 1$. Inequalities (3.63) lead to condition **(C.3)** for $\overline{D}_n^M(u)$. To get **(C.3)** for the difference $D_n^M(u) - \overline{D}_n^M(u)$, write

$$|D_n^M(u) - \overline{D}_n^M(u)| \leq \frac{\sqrt{2}}{\pi}\sqrt{n} \sum_{k=M+1}^{n-(P+1)} \|\boldsymbol{\eta}'_k(\mathbf{Z}'_k \mathcal{W}^{-1}\mathbf{Z}_k)^{-1}\| \|\mathbf{U}_k - \overline{\mathbf{U}}_k\| \, \frac{\sin(k\pi u)}{k} \ .$$

Since $\sqrt{n}(\mathbf{C}_0 - \boldsymbol{\Sigma}) = O_P(1)$, using similar techniques to the ones used in the proof of Proposition 2.6.3, it can be established that

$$\limsup_M [\limsup_n \Pr(\sup_{0\leq u\leq 1} |D_n^M(u) - \overline{D}_n^M(u)| > \varepsilon)] = 0 \ . \ \blacksquare$$

# Chapter 4

# A goodness-of-fit process based on a transformed correlation sequence

**Summary.** This chapter introduces a modified sequence of adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$, where $\widehat{\mathbf{S}}_k$ is a properly constructed $m \times m$ matrix, $k = p+q+1, \ldots, n-(P+1)$. In some sense, the $\{\widehat{\mathbf{S}}_k\}$ are obtained after transforming the original residual correlation matrices $\{\widehat{\mathbf{R}}_k : 1 \leq k \leq n-(P+1)\}$ of (1.21) (Chitturi, 1974). This leads to consider, as a replacement of the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (1.28), the modified process $\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$ of (1.30). Sections 4.1 and 4.2 establish the notation, and give some background and motivation. Section 4.3 analyzes in detail the properties of the sequence $\{\widehat{\mathbf{S}}_k\}$. The particular case of a *VAR(1)* model is described in some detail, because it serves as an illustration of the numerical strategy underlying the construction of the $\{\widehat{\mathbf{S}}_k\}$. Section 4.4 establishes convergence of $\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$ to the Brownian bridge. Section 4.5 contains some concluding remarks.

## 4.1 Introduction

For *ARMA(p,q)* models, Ubierna and Velilla (2007) consider a modified goodness-of-fit process $\{\widehat{Z}_n(u) : 0 \leq u \leq 1\}$, where

$$\widehat{Z}_n(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=p+q+1}^{n-(P+1)} \widehat{s}_k \frac{\sin(K\pi u)}{K} \; ; \tag{4.1}$$

$K = k-(p+q)$; and $\{\widehat{s}_k : p+q+1 \leq k \leq n-(P+1)\}$ is a transformed residual autocorrelation sequence. As defined in Ubierna and Velilla (2007, section 3.2),

$$\widehat{s}_k = \widehat{\boldsymbol{\gamma}}_k' \begin{pmatrix} \widehat{r}_1 \\ \widehat{r}_2 \\ \vdots \\ \widehat{r}_k \end{pmatrix} \quad , \quad p+q+1 \leq k \leq n-(P+1) \; , \tag{4.2}$$

where the $\widehat{r}_k$ are the univariate residual correlations of (1.5); $\widehat{\boldsymbol{\gamma}}_k$ is the empirical version of a $k \times 1$ unit vector $\boldsymbol{\gamma}_k$ such that $\boldsymbol{\gamma}_k'\mathbf{A}_k = \mathbf{0}$, where

$$\mathbf{A}_k = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_1 & 1 & \cdots & 0 \\ a_2 & a_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{k-1} & a_{k-2} & \cdots & a_{k-(p+q)} \end{pmatrix}$$

is the $k \times (p+q)$ matrix of (2.32); and the $\{a_r : r \geq 0\}$ are the coefficients of the series expansion $a(z) = [\phi(z)\theta(z)]^{-1} = [\theta(z)\phi(z)]^{-1} = \sum_{r=0}^{\infty} a_r z^r$. The vectors $\{\boldsymbol{\gamma}_k\}$ are constructed recursively, and for $j,k \leq M$ they satisfy the condition $\gamma^{*\prime}_j \gamma^*_k = \delta_{jk}$, where $\delta_{jk}$ is Dirac's delta, and $\boldsymbol{\gamma}^*_k = [\boldsymbol{\gamma}'_k \mid \mathbf{0}'_{(M-k)\times 1}]'$ is a $M \times 1$ unit vector. Thus, the $M \times [M-(p+q)]$ matrix with ladder structure,

$$\boldsymbol{\psi}_M = (\boldsymbol{\gamma}^*_{p+q+1} \mid \cdots \mid \boldsymbol{\gamma}^*_M) \ , \tag{4.3}$$

is such that

$$\boldsymbol{\psi}'_M \boldsymbol{\psi}_M = \mathbf{I}_{M-(p+q)} \quad , \quad \boldsymbol{\psi}'_M \mathbf{A}_M = \mathbf{0} \ . \tag{4.4}$$

The estimated version of (4.3) will be denoted $\widehat{\boldsymbol{\psi}}_M$.

The idea underlying the construction of (4.2) is as follows. Consider expression (2.31) by Box and Pierce (1970, section 5),

$$\begin{pmatrix} \widehat{r}_1 \\ \widehat{r}_2 \\ \vdots \\ \widehat{r}_M \end{pmatrix} = [\mathbf{I}_M - \mathbf{A}_M(\mathbf{A}'_M\mathbf{A}_M)^{-1}\mathbf{A}'_M] \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_M \end{pmatrix} + O_P(\frac{1}{n}) \ .$$

For $n$ large, the distribution of the random vector $(r_1, \ldots, r_M)'$ is approximately $N_M(\mathbf{0}, n^{-1}\,\mathbf{I}_M)$. Hence, from (2.31) and (4.2)$-$(4.3)$-$(4.4) it is obtained that

$$\sqrt{n} \begin{pmatrix} \widehat{s}_{p+q+1} \\ \widehat{s}_{p+q+2} \\ \vdots \\ \widehat{s}_M \end{pmatrix} = \sqrt{n}\,\widehat{\boldsymbol{\psi}}'_M \begin{pmatrix} \widehat{r}_1 \\ \widehat{r}_2 \\ \vdots \\ \widehat{r}_M \end{pmatrix} \overset{D}{\cong}$$

$$\overset{D}{\cong} \boldsymbol{\psi}'_M[\mathbf{I}_M - \mathbf{A}_M(\mathbf{A}'_M\mathbf{A}_M)^{-1}\mathbf{A}'_M]\sqrt{n} \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_M \end{pmatrix} \overset{D}{=}$$

$$\overset{D}{=} \boldsymbol{\psi}'_M\sqrt{n} \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_M \end{pmatrix} \overset{D}{\cong} N_M[\mathbf{0}, \boldsymbol{\psi}'_M\boldsymbol{\psi}_M = \mathbf{I}_{M-(p+q)}] \ . \tag{4.5}$$

From (4.5) it follows that

$$\sqrt{n} \begin{pmatrix} \widehat{s}_{p+q+1} \\ \widehat{s}_{p+q+2} \\ \vdots \\ \widehat{s}_M \end{pmatrix} \overset{D}{\cong} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{M-(p+q)} \end{pmatrix} \ , \tag{4.6}$$

where $\{v_k : k \geq 1\}$ is a sequence of i.i.d. $N(0,1)$ random variables. As a consequence of (4.6), the partial sums of the process $\{\widehat{Z}_n(u) : 0 \leq u \leq 1\}$ of (4.1) can be conjectured to have, for $n$ large, a behavior close to those of the Karhunen-Loève representation (2.49) of the Brownian bridge,

$$B(u) = \frac{\sqrt{2}}{\pi} \sum_{k=1}^{\infty} v_k \frac{\sin(k\pi u)}{k} \ , \quad 0 \leq u \leq 1 \ .$$

In fact, Ubierna and Velilla (2007, Theorem 3.1) establish that $\widehat{Z}_n(u) \to_w B(u)$.

In summary, the use of the transformation (4.2)−(4.3)−(4.4) removes the dependence on unknown parameters of the covariance function of (3.1),

$$\gamma(u,v) = [\min(u,v) - uv] - \frac{1}{2\pi^2} \mathbf{g}(\pi u)' \mathbf{I}^{-1}(\boldsymbol{\lambda}) \mathbf{g}(\pi v) \ , \qquad 0 \leq u, v \leq 1 \ ,$$

that corresponds to the limit of the process $\{\widehat{W}_n(u) : 0 \leq u \leq 1\}$, where

$$\widehat{W}_n(u) = \frac{\sqrt{2}}{\pi} \sqrt{n} \sum_{k=1}^{n-(P+1)} \widehat{r}_k \frac{\sin(k\pi u)}{k}$$

is as given in (1.25). As explained by Ubierna and Velilla (2007, section 3.1), this parametric dependence is due to the presence of the projection matrix $\mathbf{A}_M(\mathbf{A}'_M\mathbf{A}_M)^{-1}$ $\mathbf{A}'_M$ at the right-hand side of expression (2.31) of Box and Pierce (1970, section 5). From (4.5)−(4.6), the orthogonality condition $\boldsymbol{\psi}'_M\mathbf{A}_M = \mathbf{0}$ of (4.4) eliminates this projection matrix from the approximate distribution of the $\widehat{s}_k$ in (4.2). As a consequence, by Ubierna and Velilla (2007, Theorem 3.1), $\gamma(u,v)$ in (3.1) becomes $\min(u,v) - uv$, the covariance function of the Brownian bridge.

The goal of this chapter is to propose an adequate generalization of the process $\{\widehat{Z}_n(u) : 0 \leq u \leq 1\}$ of (4.1) for *VARMA(p,q)* models. A new modified goodness-of-fit process $\{\widehat{Z}^m_n(u) : 0 \leq u \leq 1\}$ is introduced, where

$$\widehat{Z}^m_n(u) = \frac{\sqrt{2}}{\pi} \sqrt{n} \sum_{k=p+q+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}} \frac{\sin(K\pi u)}{K}$$

depends, as seen in (1.30), on a modified residual matrix autocorrelation sequence $\{\widehat{\mathbf{S}}_k\}$. Definitions of the elements involved in this construction are given next.

## 4.2 Motivation

In the multivariate case, the analogue of (2.31) by Box and Pierce (1970, section 5) is given by expressions (2.29)−(2.30) (Hosking, 1980),

$$\widehat{\mathbf{W}}^{-1/2}\begin{pmatrix}\mathrm{vec}(\widehat{\mathbf{C}}_1')\\ \mathrm{vec}(\widehat{\mathbf{C}}_2')\\ \vdots\\ \mathrm{vec}(\widehat{\mathbf{C}}_M')\end{pmatrix} \overset{D}{\cong} (\mathbf{I}_{Mm^2}-\mathbf{P}_M)N_{Mm^2}(\mathbf{0},\mathbf{I}_{Mm^2}) \ ,$$

where $\mathbf{P}_M = \mathcal{W}^{-1/2}\mathbf{Z}_M(\mathbf{Z}_M'\mathcal{W}^{-1}\mathbf{Z}_M)^{-1}\mathbf{Z}_M'\mathcal{W}^{-1/2}$ is the $Mm^2 \times Mm^2$ orthogonal projection matrix onto the subspace spanned by the columns of $\mathcal{W}^{-1/2}\mathbf{Z}_M$, and $\widehat{\mathbf{W}} = \mathbf{I}_M \otimes \widehat{\mathbf{\Sigma}} \otimes \widehat{\mathbf{\Sigma}}$. Alternatively,

$$\sqrt{n}\begin{pmatrix}\mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_1'\widehat{\mathbf{\Sigma}}^{-1/2})\\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_2'\widehat{\mathbf{\Sigma}}^{-1/2})\\ \vdots\\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_M'\widehat{\mathbf{\Sigma}}^{-1/2})\end{pmatrix} \overset{D}{\cong} (\mathbf{I}_{Mm^2}-\mathbf{P}_M)N_{Mm^2}(\mathbf{0},\mathbf{I}_{Mm^2}) \ . \tag{4.7}$$

Equation (4.7) suggests that, rather than considering the residual goodness-of-fit process $\{\widehat{W}_n^m(u) : 0 \le u \le 1\}$ of (1.28), that is based on the adjusted residual traces

$$\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m} = \mathbf{a}_m'\mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_k'\widehat{\mathbf{\Sigma}}^{-1/2}) \ , \qquad 1 \le k \le n-(P+1) \ ,$$

of (1.29)−(2.36), where $\mathbf{a}_m = \mathrm{vec}(\mathbf{I_m})/\sqrt{m}$ is of $m^2 \times 1$, it may be better to construct a sequence of modified correlation $m \times m$ matrices $\{\widehat{\mathbf{S}}_k\}$ such that

$$\begin{pmatrix}\mathrm{vec}(\widehat{\mathbf{S}}_{p+q+1})\\ \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+2})\\ \vdots\\ \mathrm{vec}(\widehat{\mathbf{S}}_M)\end{pmatrix} = \widehat{\mathbf{\Psi}}_M'\begin{pmatrix}\mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_1'\widehat{\mathbf{\Sigma}}^{-1/2})\\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_2'\widehat{\mathbf{\Sigma}}^{-1/2})\\ \vdots\\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_M'\widehat{\mathbf{\Sigma}}^{-1/2})\end{pmatrix} \ , \tag{4.8}$$

where $\widehat{\mathbf{\Psi}}_M$ is the estimated version of a $Mm^2 \times m^2[M-(p+q)]$ matrix $\mathbf{\Psi}_M$ such that

$$\mathbf{\Psi}_M'\mathbf{\Psi}_M = \mathbf{I}_{m^2[M-(p+q)]} \quad , \quad \mathbf{\Psi}_M'\mathcal{W}^{-1/2}\mathbf{Z}_M = \mathbf{0} \ . \tag{4.9}$$

Combining (4.7)−(4.8)−(4.9) leads to

$$\sqrt{n}\begin{pmatrix}\mathrm{vec}(\widehat{\mathbf{S}}_{p+q+1})\\ \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+2})\\ \vdots\\ \mathrm{vec}(\widehat{\mathbf{S}}_M)\end{pmatrix} \overset{D}{\cong}$$

$$\stackrel{D}{\cong} \mathbf{\Psi}'_M(\mathbf{I}_{Mm^2} - \mathbf{P}_M)N_{Mm^2}(\mathbf{0}, \mathbf{I}_{Mm^2}) = N_{m^2[M-(p+q)]}(\mathbf{0}, \mathbf{I}_{m^2[M-(p+q)]}) \ . \tag{4.10}$$

The results in (4.8)−(4.9)−(4.10) are multivariate generalizations of the univariate expressions (4.2)−(4.3)−(4.4)−(4.5).

Define now the sequence of modified adjusted residual traces

$$\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m} = \mathbf{a}'_m \mathrm{vec}(\widehat{\mathbf{S}}_k) \ , \qquad p+q+1 \leq k \leq n-(P+1) \ . \tag{4.11}$$

Since $[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m][\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}_m] = \mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m\mathbf{a}_m = \mathbf{I}_{M-(p+q)} \otimes 1 = \mathbf{I}_{M-(p+q)}$, from (4.10) it follows that

$$\sqrt{n}\,[\frac{1}{\sqrt{m}} \begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{S}}_{p+q+1}) \\ \mathrm{tr}(\widehat{\mathbf{S}}_{p+q+2}) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{S}}_M) \end{pmatrix}] =$$

$$= \sqrt{n}\,[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m] \begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+1}) \\ \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+2}) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{S}}_M) \end{pmatrix} \stackrel{D}{\cong} N_{M-(p+q)}[\mathbf{0}, \mathbf{I}_{M-(p+q)}] \ . \tag{4.12}$$

As a conclusion from (4.12),

$$\sqrt{n}\,[\frac{1}{\sqrt{m}} \begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{S}}_{p+q+1}) \\ \mathrm{tr}(\widehat{\mathbf{S}}_{p+q+2}) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{S}}_M) \end{pmatrix}] \stackrel{D}{\cong} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{M-(p+q)} \end{pmatrix} \ , \tag{4.13}$$

where $\{v_k : k \geq 1\}$ is a sequence of i.i.d. $N(0,1)$ random variables.

The approximation of (4.13) for the modified adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ of (4.11) is similar to that obtained in (4.6) for the transformed residual autocorrelations $\widehat{s}_k$ of (4.2). Thus, in agreement with the comments made in section 4.1, it may be conjectured that the process $\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$ of (1.30), where

$$\widehat{Z}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n} \sum_{k=p+q+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}} \frac{\sin(K\pi u)}{K} \ ,$$

will converge to the Brownian bridge. This asymptotic behavior is an improvement over the parametric limit properties of the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (1.28), that are described in Theorem 3.4.1.

Before going into details, it is necessary to specify first the explicit expression of the matrices $\mathbf{\Psi}_M$ that appear in (4.8)−(4.9). This is studied next.

## 4.3 Construction and properties of the sequence of transformation matrices

### 4.3.1 Discussion of the univariate case

Equations (4.9) and (4.12) are multivariate analogues of the univariate expressions (4.4) and (4.5). Thus, it may be useful to describe first the construction of the sequence of $k \times 1$ unit vectors $\{\boldsymbol{\gamma}_k : p+q+1 \leq k \leq n-(P+1)\}$ that lead to the matrix $\boldsymbol{\psi}_M$ of (4.3). Ubierna and Velilla (2007, section 3.3) suggested taking

$$\boldsymbol{\gamma}_k = \frac{1}{\sqrt{1+\boldsymbol{\zeta}_k'(\mathbf{A}_{k-1}'\mathbf{A}_{k-1})^{-1}\boldsymbol{\zeta}_k}} \begin{pmatrix} -\mathbf{A}_{k-1}(\mathbf{A}_{k-1}'\mathbf{A}_{k-1})^{-1}\boldsymbol{\zeta}_k \\ 1 \end{pmatrix} , \tag{4.14}$$

$p+q+1 \leq k \leq n-(P+1)$, where $\mathbf{A}_{k-1}$ is the $(k-1)\times(p+q)$ matrix of (2.32); and

$$\boldsymbol{\zeta}_k' = [a_{k-1}, a_{k-2}, \ldots, a_{k-(p+q)}] \tag{4.15}$$

is the $k$th $1\times(p+q)$ row vector of the matrix $\mathbf{A}_k$. Construction (4.14)$-$(4.15) above is motivated by Velilla (1994, section 3.2).

From (4.14)$-$(4.15) it is easy to see that $\boldsymbol{\gamma}_k'\mathbf{A}_k = \mathbf{0}$, and $\gamma^{*\prime}_j\gamma^*_k = \delta_{jk}$ for $j,k \leq M$, where $\boldsymbol{\gamma}^*_k = [\boldsymbol{\gamma}_k' \mid \mathbf{0}'_{(M-k)\times 1}]'$ is of $M \times 1$. Thus, the $M \times [M-(p+q)]$ matrix

$$\boldsymbol{\psi}_M = (\boldsymbol{\gamma}^*_{p+q+1} \mid \cdots \mid \boldsymbol{\gamma}^*_M)$$

of (4.3) is an orthonormal basis of the orthogonal complement in $\mathbb{R}^M$ of the column space of $\mathbf{A}_M$ in (2.32). The sample $k \times 1$ vector $\widehat{\boldsymbol{\gamma}}_k$ that appears in the modified correlation $\widehat{s}_k$ of (4.2) uses in expression (4.14) the coefficients of the expansion $\widehat{a}(z) = [\widehat{\phi}(z)\widehat{\theta}(z)]^{-1} = [\widehat{\theta}(z)\widehat{\phi}(z)]^{-1} = \sum_{r=0}^{\infty}\widehat{a}_r z^r$, associated to the estimates $(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}})$.

Recall now that $|a_k| \leq ab^k$, $k \geq 0$, where $a > 0$, and $0 < b < 1$. Thus, taking the limit in (4.14) as $k \to \infty$ it follows that

$$\boldsymbol{\gamma}_k \to \mathbf{e}_k \ , \tag{4.16}$$

where $\mathbf{e}_k$ is the $k$th canonical vector of $\mathbb{R}^k$. From expression (4.16); consistency of $(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}})$ for $(\boldsymbol{\phi}, \boldsymbol{\theta})$; and definition (4.2), for $k$ and $n$ large enough the modified $\widehat{s}_k$ and the original $\widehat{r}_k$ will be close to each other. This is because

$$\widehat{s}_k = \widehat{\boldsymbol{\gamma}}_k' \begin{pmatrix} \widehat{r}_1 \\ \widehat{r}_2 \\ \vdots \\ \widehat{r}_k \end{pmatrix} \cong \boldsymbol{\gamma}_k' \begin{pmatrix} \widehat{r}_1 \\ \widehat{r}_2 \\ \vdots \\ \widehat{r}_k \end{pmatrix} \cong \mathbf{e}_k' \begin{pmatrix} \widehat{r}_1 \\ \widehat{r}_2 \\ \vdots \\ \widehat{r}_k \end{pmatrix} = \widehat{r}_k \ . \tag{4.17}$$

As a consequence of (4.17), it is only necessary to construct the first transformed residual autocorrelations $\widehat{s}_k$ of (4.2) for $p+q+1 \leq k \leq M$, for a suitable chosen lag $M$. Then, for practical purposes the components of the process $\{\widehat{Z}_n(u) : 0 \leq u \leq 1\}$ of (4.1) can be approximately written in the form

$$\widehat{Z}_n(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=p+q+1}^{n-(P+1)} \widehat{s}_k \frac{\sin(K\pi u)}{K} \cong$$

$$\cong \frac{\sqrt{2}}{\pi}\sqrt{n}\,[\sum_{k=p+q+1}^{M} \widehat{s}_k \frac{\sin(K\pi u)}{K} + \sum_{k=M+1}^{n-(P+1)} \widehat{r}_k \frac{\sin(K\pi u)}{K}] \ , \tag{4.18}$$

where $K = k-(p+q)$. The structure of (4.18) is similar to that of

$$\widehat{W}_n(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-(P+1)} \widehat{r}_k \frac{\sin(k\pi u)}{k} =$$

$$\frac{\sqrt{2}}{\pi}\sqrt{n}\,[\sum_{k=1}^{M} \widehat{r}_k \frac{\sin(k\pi u)}{k} + \sum_{k=M+1}^{n-(P+1)} \widehat{r}_k \frac{\sin(k\pi u)}{k}]$$

in (1.25). By considering the vectors $\boldsymbol{\alpha}_M(u)$ of (2.50) and definition (4.2)$-$(4.3)$-$(4.4) the first summands of (4.18) and (1.25) can be written in the form

$$\boldsymbol{\alpha}'_{M-(p+q)}(u)\sqrt{n}\,\widehat{\boldsymbol{\psi}}'_M \begin{pmatrix} \widehat{r}_1 \\ \widehat{r}_2 \\ \vdots \\ \widehat{r}_M \end{pmatrix} \quad , \quad \boldsymbol{\alpha}'_M(u)\sqrt{n} \begin{pmatrix} \widehat{r}_1 \\ \widehat{r}_2 \\ \vdots \\ \widehat{r}_M \end{pmatrix} . \tag{4.19}$$

Both expressions in (4.19) depend on the leading $M$ residual autocorrelations $\widehat{r}_k$, $1 \leq k \leq M$, of (1.5). However, in the term at the left these statistics are processed by the estimated $M \times [M-(p+q)]$ ladder matrix $\widehat{\boldsymbol{\psi}}_M$ of (4.3).

On the other hand, as mentioned in section 2.4.2, it can be written

$$\mathbf{A}_M = \mathbf{X}_M \mathbf{B} \ , \tag{4.20}$$

where $\mathbf{B}$ is an invertible $(p+q) \times (p+q)$ matrix, and

$$\mathbf{X}_M = \left(\begin{array}{cccc|cccc}
1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
h_1 & 1 & \cdots & \vdots & l_1 & 1 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\
h_{p-1} & h_{p-2} & \cdots & 1 & \vdots & \vdots & \ddots & \vdots \\
\cline{1-4}
h_p & h_{p-1} & \cdots & h_1 & \vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
\vdots & \vdots & \ddots & \vdots & l_{q-1} & l_{q-2} & \cdots & 1 \\
\cline{5-8}
\vdots & \vdots & \ddots & \vdots & l_q & l_{q-1} & \cdots & l_1 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\
h_{M-1} & h_{M-2} & \cdots & h_{M-p} & l_{M-1} & l_{M-2} & \cdots & l_{M-q}
\end{array}\right)$$

is the $M \times (p+q)$ matrix considered by McLeod (1979) in expression (1.8). The entries of $\mathbf{X}_M$ are given by the coefficients $\{h_r : r \geq 0\}$ of the series $\phi^{-1}(z) = \sum_{r=0}^{\infty} h_r z^r$, where $h_0 = 1$; and the coefficients $\{l_r : r \geq 0\}$ of the series $\theta^{-1}(z) = \sum_{r=0}^{\infty} l_r z^r$, where $l_0 = 1$. Equation (4.20) follows from the commutativity property $\phi(B)\theta(B) = \theta(B)\phi(B)$, that allows to write $l_k = \phi(B)a_k$; and $h_k = \theta(B)a_k$. Details, that are algebraic in nature, are omitted for conciseness.

A consequence of identity (4.20) is that the transformation vectors of definition (4.14) can be also written in the form

$$\boldsymbol{\gamma}_k = \frac{1}{\sqrt{1 + \boldsymbol{\xi}_k'(\mathbf{X}_{k-1}'\mathbf{X}_{k-1})^{-1}\boldsymbol{\xi}_k}} \begin{pmatrix} -\mathbf{X}_{k-1}(\mathbf{X}_{k-1}'\mathbf{X}_{k-1})^{-1}\boldsymbol{\xi}_k \\ 1 \end{pmatrix} , \tag{4.21}$$

$p + q + 1 \leq k \leq n - (P + 1)$, where

$$\boldsymbol{\xi}_k' = (h_{k-1}, h_{k-2}, \ldots, h_{k-p}; l_{k-1}, l_{k-2}, \ldots, l_{k-q}) \tag{4.22}$$

is the $k$th row of the matrix $\mathbf{X}_M$ by McLeod (1979) in (1.8). Re-expression (4.21) for the $\boldsymbol{\gamma}_k$ in (4.14) is useful for inspiring the definition of the transformation matrices $\boldsymbol{\Psi}_M$ in (4.9). This is because, by expression (4.22), the $m^2 \times m^2(p+q)$ row blocks of the matrix $\mathcal{W}^{-1/2}\mathbf{Z}_M$ given in (2.23),

$$\boldsymbol{\Xi}_k' = (\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2})(\mathbf{G}_{k-1}, \ldots, \mathbf{G}_{k-p}\,; \mathbf{F}_{k-1}, \ldots, \mathbf{F}_{k-q}) \ ,$$

have an structure similar to that of the rows $\boldsymbol{\xi}_k'$ of $\mathbf{X}_M$ in (1.8). Moreover, from section 2.4.2, for $m = 1$ it follows that $\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2} = 1/\sigma^2$; $\mathbf{G}_k = \sigma^2 h_k$; and $\mathbf{F}_k = \sigma^2 l_k$.

Thus, $\boldsymbol{\Xi}'_k = \boldsymbol{\xi}'_k$. On the other hand, by Hosking (1980, section 8) and Pierce (1970), when $m > 1$ there is not an analogue to identity (4.20) for the matrix $\mathcal{W}^{-1/2}\mathbf{Z}_M$.

Other formulations for the sequence of unit transformation vectors vectors $\{\boldsymbol{\gamma}_k : p+q+1 \leq k \leq n-(P+1)\}$ could be possible. However, the explicit construction (4.14)−(4.21) is a convenient device for both asymptotic and numerical purposes.

### 4.3.2 The multivariate case

This section contains a proposal for the transformation matrices $\boldsymbol{\Psi}_M$ of (4.9) that generalizes, in agreement with the guidelines of the previous section, the structure of the univariate matrices $\boldsymbol{\psi}_M$ of expressions (4.3)−(4.4).

Consider, for each $p+q+1 \leq k \leq n-(P+1)$, the $km^2 \times m^2$ matrix

$$\boldsymbol{\Gamma}_k = \begin{pmatrix} -\mathcal{W}^{-1/2}\mathbf{Z}_{k-1}(\mathbf{Z}'_{k-1}\mathcal{W}^{-1}\mathbf{Z}_{k-1})^{-1}\boldsymbol{\Xi}_k \\ \mathbf{I}_{m^2} \end{pmatrix} [\mathbf{I}_{m^2} + \boldsymbol{\Xi}'_k(\mathbf{Z}'_{k-1}\mathcal{W}^{-1}\mathbf{Z}_{k-1})^{-1}\boldsymbol{\Xi}_k]^{-1/2} , \tag{4.23}$$

where $\mathcal{W} = \mathbf{I}_k \otimes \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}$. The structure of (4.23) can be seen as a multivariate version of expression (4.21), with the matrix $\mathbf{X}_{k-1}$ of (1.8) replaced by $\mathcal{W}^{-1/2}\mathbf{Z}_{k-1}$; and the row $\boldsymbol{\xi}'_k$ of (4.22) by the row block $\boldsymbol{\Xi}'_k$ in (2.23). To simplify the notation, in what follows it is convenient to write, for $p+q+1 \leq k \leq n-(P+1)$,

$$\boldsymbol{\Gamma}_k = \boldsymbol{\Delta}_k(\boldsymbol{\Delta}'_k\boldsymbol{\Delta}_k)^{-1/2} , \tag{4.24}$$

where

$$\boldsymbol{\Delta}_k = \begin{pmatrix} -\mathcal{W}^{-1/2}\mathbf{Z}_{k-1}(\mathbf{Z}'_{k-1}\mathcal{W}^{-1}\mathbf{Z}_{k-1})^{-1}\boldsymbol{\Xi}_k \\ \mathbf{I}_{m^2} \end{pmatrix} \tag{4.25}$$

is of $km^2 \times m^2$, and $\boldsymbol{\Delta}'_k\boldsymbol{\Delta}_k = \mathbf{I}_{m^2} + \boldsymbol{\Xi}'_k(\mathbf{Z}'_{k-1}\mathcal{W}^{-1}\mathbf{Z}_{k-1})^{-1}\boldsymbol{\Xi}_k > \mathbf{0}$.

The $Mm^2 \times m^2[M-(p+q)]$ matrix $\boldsymbol{\Psi}_M$ of (4.9) is constructed by considering the collection of $Mm^2 \times m^2$ matrices

$$\boldsymbol{\Gamma}^*_k = \left( \frac{\boldsymbol{\Gamma}_k}{\mathbf{0}_{(M-k)m^2 \times m^2}} \right) , \quad k = p+q+1, \ldots, M , \tag{4.26}$$

where $\mathbf{0}_{(M-k)m^2 \times m^2}$ is a zero matrix; and defining

$$\boldsymbol{\Psi}_M = (\boldsymbol{\Gamma}^*_{p+q+1} \,|\, \boldsymbol{\Gamma}^*_{p+q+2} \,|\cdots|\, \boldsymbol{\Gamma}^*_M) . \tag{4.27}$$

By expressions (5.31)−(4.27), $\boldsymbol{\Psi}_M$ has a ladder structure. Next result formalizes conditions (4.9) for $\boldsymbol{\Psi}_M$ in (4.23)−(5.31)−(4.27).

**Proposition 4.3.1** As defined by expressions (4.23)–(5.31)–(4.27), the $Mm^2 \times m^2[M-(p+q)]$ matrix $\mathbf{\Psi}_M$ satisfies:

(a) $\mathbf{\Psi}'_M \mathbf{\Psi}_M = \mathbf{I}_{m^2[M-(p+q)]}$ .

(b) $\mathbf{\Psi}'_M \mathcal{W}^{-1/2} \mathbf{Z}_M = \mathbf{0}$ .

**Proof.** (a) It is enough to verify that $\mathbf{\Gamma}^{*\prime}_k \mathbf{\Gamma}^*_k = \mathbf{I}_{m^2}$, $k = p+q+1, \ldots, M$; and $\mathbf{\Gamma}^{*\prime}_j \mathbf{\Gamma}^*_k = \mathbf{0}$, for $j < k \leq M$. Using (4.24)–(4.25),

$$\mathbf{\Gamma}^{*\prime}_k \mathbf{\Gamma}^*_k = \mathbf{\Gamma}'_k \mathbf{\Gamma}_k = (\mathbf{\Delta}'_k \mathbf{\Delta}_k)^{-1/2} \mathbf{\Delta}'_k \mathbf{\Delta}_k (\mathbf{\Delta}'_k \mathbf{\Delta}_k)^{-1/2} = \mathbf{I}_{m^2} \ .$$

For $j < k \leq M$, the condition $\mathbf{\Gamma}^{*\prime}_j \mathbf{\Gamma}^*_k = \mathbf{0}$ follows from the representation

$$\mathbf{\Gamma}^{*\prime}_j \mathbf{\Gamma}^*_k = (\mathbf{\Delta}'_j \mathbf{\Delta}_j)^{-1/2} [\mathbf{\Delta}'_j \mid \mathbf{0}_{m^2 \times (k-j)m^2}] \mathbf{\Delta}_k (\mathbf{\Delta}'_k \mathbf{\Delta}_k)^{-1/2} \ ;$$

and the fact that, from definition (4.25), it can be written

$$[\mathbf{\Delta}'_j \mid \mathbf{0}_{m^2 \times (k-j)m^2}] \mathbf{\Delta}_k =$$

$$= [-\mathbf{\Xi}'_j (\mathbf{Z}'_{j-1} \mathcal{W}^{-1} \mathbf{Z}_{j-1})^{-1} \mathbf{Z}'_{j-1} \mathcal{W}^{-1/2} \mid \mathbf{I}_{m^2}] \begin{pmatrix} -\mathcal{W}^{-1/2} \mathbf{Z}_{j-1} \\ -\mathbf{\Xi}'_j \end{pmatrix} (\mathbf{Z}'_{k-1} \mathcal{W}^{-1} \mathbf{Z}_{k-1})^{-1} \mathbf{\Xi}_k =$$

$$= (\mathbf{\Xi}'_j - \mathbf{\Xi}'_j)(\mathbf{Z}'_{k-1} \mathcal{W}^{-1} \mathbf{Z}_{k-1})^{-1} \mathbf{\Xi}_k = \mathbf{0} \ .$$

(b) For obtaining the condition $\mathbf{\Gamma}^{*\prime}_k \mathcal{W}^{-1/2} \mathbf{Z}_M = \mathbf{0}$ for $k = p+q+1, \ldots, M$, put

$$\mathbf{\Gamma}^{*\prime}_k \mathcal{W}^{-1/2} \mathbf{Z}_M = (\mathbf{\Delta}'_k \mathbf{\Delta}_k)^{-1/2} [\mathbf{\Delta}'_k \mid \mathbf{0}_{m^2 \times (M-k)m^2}] \mathcal{W}^{-1/2} \mathbf{Z}_M \ ;$$

and use the identity

$$[\mathbf{\Delta}'_k \mid \mathbf{0}_{m^2 \times (M-k)m^2}] \mathcal{W}^{-1/2} \mathbf{Z}_M =$$

$$= [-\mathbf{\Xi}'_k (\mathbf{Z}'_{k-1} \mathcal{W}^{-1} \mathbf{Z}_{k-1})^{-1} \mathbf{Z}'_{k-1} \mathcal{W}^{-1/2} \mid \mathbf{I}_{m^2}] \begin{pmatrix} \mathcal{W}^{-1/2} \mathbf{Z}_{k-1} \\ \mathbf{\Xi}'_k \end{pmatrix} = -\mathbf{\Xi}'_k + \mathbf{\Xi}'_k = \mathbf{0} \ . \blacksquare$$

From part (b) of Proposition 4.3.1, the $Mm^2 \times m^2[M-(p+q)]$ matrix $\mathbf{\Psi}_M$ is an orthonormal basis of the orthogonal complement in $\mathbb{R}^{Mm^2}$ of the column space of the $Mm^2 \times m^2(p+q)$ matrix $\mathcal{W}^{-1/2} \mathbf{Z}_M$.

### 4.3.3 Numerical implications

According to (4.8); (4.23)−(5.31)−(4.27); and (4.24)−(4.25), the $k$th modified residual correlation $m \times m$ matrix $\widehat{\mathbf{S}}_k$ is such that

$$\mathrm{vec}(\widehat{\mathbf{S}}_k) = \widehat{\mathbf{\Gamma}}'_k \begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_1\widehat{\mathbf{\Sigma}}^{-1/2}) \\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_2\widehat{\mathbf{\Sigma}}^{-1/2}) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1/2}) \end{pmatrix} , \quad p+q+1 \le k \le n-(P+1) \,, \qquad (4.28)$$

where $\widehat{\mathbf{\Gamma}}_k = \widehat{\mathbf{\Delta}}_k(\widehat{\mathbf{\Delta}}'_k\widehat{\mathbf{\Delta}}_k)^{-1/2}$ is an empirical version of $\mathbf{\Gamma}_k = \mathbf{\Delta}_k(\mathbf{\Delta}'_k\mathbf{\Delta}_k)^{-1/2}$ in (4.24)−(4.25), constructed from suitable ML estimators $(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}}, \widehat{\mathbf{\Sigma}})$ of the parameters $(\mathbf{\Phi}, \mathbf{\Theta}, \mathbf{\Sigma})$ of model (1.1). From expression (4.28), it can be written

$$\mathrm{vec}(\widehat{\mathbf{S}}_k) = [\mathbf{I}_{m^2} + \widehat{\mathbf{\Xi}}'_k(\widehat{\mathbf{Z}}'_{k-1}\widehat{\mathcal{W}}^{-1}\widehat{\mathbf{Z}}_{k-1})^{-1}\widehat{\mathbf{\Xi}}_k]^{-1/2}$$

$$[\,\mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1/2}) - \widehat{\mathbf{\Xi}}'_k(\widehat{\mathbf{Z}}'_{k-1}\widehat{\mathcal{W}}^{-1}\widehat{\mathbf{Z}}_{k-1})^{-1}\sum_{j=1}^{k-1}\widehat{\mathbf{\Xi}}_j\mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_j\widehat{\mathbf{\Sigma}}^{-1/2})\,] \,, \qquad (4.29)$$

$p+q+1 \le k \le n-(P+1)$. The specific entries of the $m \times m$ matrix

$$\widehat{\mathbf{S}}_k = (\widehat{s}_{ij,k} : i,j = 1,\ldots,m) \,, \qquad (4.30)$$

can be obtained from (4.29) using the extraction operations

$$\widehat{s}_{ij,k} = \mathbf{e}'_i\widehat{\mathbf{S}}_k\mathbf{e}_j = \mathrm{tr}(\mathbf{e}_j\mathbf{e}'_i\widehat{\mathbf{S}}_k) = [\mathrm{vec}(\mathbf{e}_i\mathbf{e}'_j)]'\mathrm{vec}(\widehat{\mathbf{S}}_k) \,, \quad i,j = 1,\ldots,m \,.$$

According to (4.29), the array $\mathrm{vec}(\widehat{\mathbf{S}}_k)$ is obtained as a linear combination of the vectorizations of the $m\times m$ matrices $\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_j\widehat{\mathbf{\Sigma}}^{-1/2}$, $j = 1, \ldots, k$. These are related to the $m \times m$ residual autocorrelation matrices $\widehat{\mathbf{R}}_j = \widehat{\mathbf{C}}'_j\widehat{\mathbf{\Sigma}}^{-1}$, $j = 1, \ldots, k$, of expression (1.21) (Chitturi, 1974). For this reason, the $\widehat{\mathbf{S}}_k$ of (4.30) can be thought, in some sense, as obtained after transforming the first $\widehat{\mathbf{R}}_j$, $j = 1, \ldots, k$. Notice also that, by definition (1.20), when $m = 1$ both $\widehat{\mathbf{R}}_j$ and $\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_j\widehat{\mathbf{\Sigma}}^{-1/2}$ coincide with $\widehat{r}_j$, the $j$th univariate residual autocorrelation of (1.5). On the other hand, taking into account that $\widehat{\mathbf{Z}}'_k\widehat{\mathcal{W}}^{-1}\widehat{\mathbf{Z}}_k = \sum_{j=1}^{k}\widehat{\mathbf{\Xi}}_k\widehat{\mathbf{\Xi}}'_k$, expression (4.29) has a recursive structure in the terms $\widehat{\mathbf{\Xi}}_k$ and $\mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1/2})$, that may be of interest for computational purposes.

The following result is useful for studying the connection between the modified adjusted residual traces of (4.11),

$$\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m} = \mathbf{a}'_m\mathrm{vec}(\widehat{\mathbf{S}}_k) \,, \qquad p+q+1 \le k \le n-(P+1) \,,$$

where $\mathbf{a}_m = \mathrm{vec}(\mathbf{I_m})/\sqrt{m}$ is of $m^2 \times 1$, and the original adjusted residual traces of expressions (1.29)−(2.36) of section 1,

$$\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m} = \mathbf{a}'_m \mathrm{vec}(\widehat{\mathbf{R}}_k) = \mathbf{a}'_m \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1/2}) \ , \quad 1 \leq k \leq n-(P+1) \ ,$$

where $\widehat{\mathbf{R}}_k = \widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1}$ is the matrix (1.21) of Chitturi (1974).

**Lemma 4.3.1** As $k \to \infty$,

$$\mathbf{\Gamma}_k = \mathbf{\Delta}_k(\mathbf{\Delta}'_k\mathbf{\Delta}_k)^{-1/2} \to \mathbf{E}_k = \left(\frac{\mathbf{0}_{\,(k-1)m^2 \times m^2}}{\mathbf{I}_{m^2}}\right) \ . \tag{4.31}$$

**Proof.** Convergence (4.31) follows from the bounds for the components of

$$\mathbf{\Xi}'_k = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})(\mathbf{G}_{k-1}, \ldots, \mathbf{G}_{k-p}\,; \mathbf{F}_{k-1}, \ldots, \mathbf{F}_{k-q}) \ ,$$

that are given in expression (3.40) of Appendix 3.1; the limit representation (2.24) for the information matrix $\mathbf{I}(\mathbf{\Lambda})$; and expression (4.25) for $\mathbf{\Delta}_k$. ∎

From convergence (4.31) in Lemma 4.3.1 and consistency of $\widehat{\mathbf{\Gamma}}_k$ to $\mathbf{\Gamma}_k$, it follows that for $k$ and $n$ large enough $\widehat{\mathbf{\Gamma}}'_k \cong (\mathbf{0}_{m^2 \times (k-1)m^2} \mid \mathbf{I}_{m^2})$. Hence, expression (4.28) simplifies to $\mathrm{vec}(\widehat{\mathbf{S}}_k) \cong \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1/2})$. Thus, using the first identity in (2.36),

$$\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m} = \mathbf{a}'_m \mathrm{vec}(\widehat{\mathbf{S}}_k) \cong \mathbf{a}'_m \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1/2}) = \mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m} \ . \tag{4.32}$$

As a consequence of (4.32), for $k$ and $n$ large enough the modified adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ of (4.11) will be close to the original statistics $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ of (1.29). This proximity is analogue to the one analyzed in section 4.3.1 that exists between the univariate statistics $\widehat{s}_k$ of (4.2), and the residual autocorrelations $\widehat{r}_k$ of (1.5).

In conclusion, the functions of the modified process $\{\widehat{Z}^m_n(u) : 0 \leq u \leq 1\}$ of (1.30) can be approximated by

$$\widehat{Z}^m_n(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=p+q+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}\frac{\sin(K\pi u)}{K} \cong$$

$$\cong \frac{\sqrt{2}}{\pi}\sqrt{n}\,[\sum_{k=p+q+1}^{M} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}\frac{\sin(K\pi u)}{K} + \sum_{k=M+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(K\pi u)}{K}] \ , \tag{4.33}$$

for a suitable selection of the lag $M$. The structure of (4.33) can be compared to that of $\widehat{W}^m_n(u)$ in (1.28), where

$$\widehat{W}^m_n(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k} =$$

$$= \frac{\sqrt{2}}{\pi}\sqrt{n}\,[\sum_{k=1}^{M}\frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k} + \sum_{k=M+1}^{n-(P+1)}\frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k}] \ .$$

Consider the vectors $\boldsymbol{\alpha}_M(u)$ of (2.50). Using identities (4.8)$-$(4.12), the first summand of (4.33) can be written

$$\boldsymbol{\alpha}'_{M-(p+q)}(u)[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m]\sqrt{n}\,\widehat{\boldsymbol{\Psi}}'_M \begin{pmatrix} \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_1\widehat{\boldsymbol{\Sigma}}^{-1/2}) \\ \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_2\widehat{\boldsymbol{\Sigma}}^{-1/2}) \\ \vdots \\ \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_M\widehat{\boldsymbol{\Sigma}}^{-1/2}) \end{pmatrix} \ . \tag{4.34}$$

In turn, using the adjusted residual traces of expressions (1.29)$-$(2.36), $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m} = \mathbf{a}'_m\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_k\widehat{\boldsymbol{\Sigma}}^{-1/2}), 1 \le k \le n-(P+1)$, the first summand of (1.28) is

$$\boldsymbol{\alpha}'_M(u)(\mathbf{I}_M \otimes \mathbf{a}'_m)\sqrt{n} \begin{pmatrix} \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_1\widehat{\boldsymbol{\Sigma}}^{-1/2}) \\ \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_2\widehat{\boldsymbol{\Sigma}}^{-1/2}) \\ \vdots \\ \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_M\widehat{\boldsymbol{\Sigma}}^{-1/2}) \end{pmatrix} \ . \tag{4.35}$$

Representations (4.34) and (4.35) are suitable multivariate generalizations of the univariate results of (4.19). Both expressions depend on the vectorizations of the matrices $\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_k\widehat{\boldsymbol{\Sigma}}^{-1/2}$, $k = 1, \ldots, M$, either directly, as in (4.35); or multiplied on the left by the transpose of the estimated matrix $\widehat{\boldsymbol{\Psi}}_M$ of (4.27), as in (4.34).

### 4.3.4 An example: the *VAR(1)* model

In general, the components of the transformation matrix $\boldsymbol{\Psi}_M$ of (4.27) must be determined numerically. Some explicit expressions can be found for a *VAR(1)* model

$$\mathbf{X}_t - \boldsymbol{\Phi}_1\mathbf{X}_{t-1} = \boldsymbol{\varepsilon}_t \ ,$$

where $\boldsymbol{\Phi}_1$ is a $m \times m$ matrix with eigenvalues $\delta_j$ such that $0 < |\delta_j| < 1$, $j = 1, \ldots, m$. In this case, $p = 1$, $q = 0$, and thus $P = \max(p, q) = 1$. Also, $\boldsymbol{\Phi}^{-1}(z)\boldsymbol{\Theta}(z) = \boldsymbol{\Phi}^{-1}(z) = \sum_{r=0}^{\infty}\boldsymbol{\Phi}_1^r z^r$. Thus, $\mathbf{G}_r = \boldsymbol{\Sigma}(\boldsymbol{\Phi}_1^r)' \otimes \mathbf{I}_m$, $r \ge 0$. Hence, the $j$th row-block (2.23) of the matrix $\mathcal{W}^{-1/2}\mathbf{Z}_k$ is of the form

$$\boldsymbol{\Xi}'_j = (\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2})\mathbf{G}_{j-1} = \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Phi}_1^{j-1})' \otimes \boldsymbol{\Sigma}^{-1/2} \ , \quad j \ge 1 \ , \tag{4.36}$$

where $\boldsymbol{\Phi}_1^0 = \mathbf{I}_m$. From (4.36) it can be written

$$\mathbf{Z}'_M\mathcal{W}^{-1}\mathbf{Z}_M = \sum_{j=1}^{M}\boldsymbol{\Xi}_j\boldsymbol{\Xi}'_j =$$

$$= \sum_{j=0}^{M-1} (\mathbf{\Phi}_1^j \mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{-1/2})[\mathbf{\Sigma}^{1/2}(\mathbf{\Phi}_1^j)' \otimes \mathbf{\Sigma}^{-1/2}] = [\sum_{j=0}^{M-1} \mathbf{\Phi}_1^j \mathbf{\Sigma}(\mathbf{\Phi}_1^j)'] \otimes \mathbf{\Sigma}^{-1} \ . \tag{4.37}$$

Expressions (4.36) and (4.37) can be used in the recursion of (4.29), after replacing the matrix $\mathbf{\Phi}_1$ by the Yule -Walker estimator $\widehat{\mathbf{\Phi}}_1$ of (2.11); and $\mathbf{\Sigma}$ by $\widehat{\mathbf{\Sigma}}$ in (2.18). This example generalizes the univariate *AR(1)* example considered in Ubierna and Velilla (2007, section 3.4).

## 4.4 Convergence to the Brownian bridge

This section formalizes the limit properties of the modified process $\{\widehat{Z}_n^m(u) : 0 \le u \le 1\}$ of expression (1.30),

$$\widehat{Z}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n} \sum_{k=p+q+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}} \frac{\sin(K\pi u)}{K} \ ,$$

where $K = k - (p+q)$.

**Theorem 4.4.1** Under the same assumptions for the errors of model (1.1) than those given in Theorem 2.6.2, as $n \to \infty$

$$\widehat{Z}_n^m(u) \to_w B(u) \ . \tag{4.38}$$

**Proof.** The technique of proof is again based on Lemma 2.6.1. The first step is to choose the proper decompositions of both

$$\widehat{Z}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n} \sum_{k=p+q+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}} \frac{\sin(K\pi u)}{K} = A_n^M(u) + R_n^M(u) \ ;$$

and

$$B(u) = \frac{\sqrt{2}}{\pi} \sum_{k=1}^{\infty} v_k \frac{\sin(K\pi u)}{K} = A^M(u) + R^M(u) \ ,$$

for $M \ge M_0$, where $M_0$ is a properly selected fixed integer number.

**First part.** For choosing $A_n^M(u)$, first notice that from (4.8) it can be written

$$\begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+1}) \\ \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+2}) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{S}}_M) \end{pmatrix} = \widehat{\mathbf{\Psi}}'_M \begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_1\widehat{\mathbf{\Sigma}}^{-1/2}) \\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_2\widehat{\mathbf{\Sigma}}^{-1/2}) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_M\widehat{\mathbf{\Sigma}}^{-1/2}) \end{pmatrix} = \widehat{\mathbf{\Psi}}'_M \widehat{\mathbf{W}}^{-1/2}\widehat{\mathbf{G}}_M \ . \tag{4.39}$$

where $\widehat{\mathbf{W}}^{-1/2} = \mathbf{I}_M \otimes \widehat{\boldsymbol{\Sigma}}^{-1/2} \otimes \widehat{\boldsymbol{\Sigma}}^{-1/2}$; $\widehat{\mathbf{G}}_M = \{[\mathrm{vec}(\widehat{\mathbf{C}}'_1)]', \ldots, [\mathrm{vec}(\widehat{\mathbf{C}}'_M)]'\}'$; and $\widehat{\boldsymbol{\Psi}}_M$ is the estimated version of the matrix $\boldsymbol{\Psi}_M$ of Proposition 4.3.1. Using (4.12) it follows from expression (4.39) that

$$\frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=p+q+1}^{M}\frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}\frac{\sin(K\pi u)}{K} =$$

$$= \quad \boldsymbol{\alpha}'_{M-(p+q)}(u)\sqrt{n}\,[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{S}}_{p+q+1}) \\ \mathrm{tr}(\widehat{\mathbf{S}}_{p+q+2}) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{S}}_M) \end{pmatrix}] \quad =$$

$$= \quad \boldsymbol{\alpha}'_{M-(p+q)}(u)\sqrt{n}\,[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m]\begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+1}) \\ \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+2}) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{S}}_M) \end{pmatrix} \quad =$$

$$= \quad \boldsymbol{\alpha}'_{M-(p+q)}(u)[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m]\widehat{\boldsymbol{\Psi}}'_M\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}\sqrt{n}\,\mathcal{W}^{-1/2}\widehat{\mathbf{G}}_M \ , \tag{4.40}$$

where $\boldsymbol{\alpha}_{M-(p+q)}(u)$ is the $[M-(p+q)] \times 1$ vector of (2.50). From decomposition (4.40), it can be written

$$\frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=p+q+1}^{M}\frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}\frac{\sin(K\pi u)}{K} =$$

$$\begin{aligned} = \quad & \boldsymbol{\alpha}'_{M-(p+q)}(u)[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m](\widehat{\boldsymbol{\Psi}}_M - \boldsymbol{\Psi}_M)'\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}\sqrt{n}\,\mathcal{W}^{-1/2}\widehat{\mathbf{G}}_M \\ & + \sqrt{n}\boldsymbol{\alpha}'_{M-(p+q)}(u)[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m]\boldsymbol{\Psi}'_M\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}\sqrt{n}\,\mathcal{W}^{-1/2}\widehat{\mathbf{G}}_M \ . \end{aligned} \tag{4.41}$$

As seen in expression (3.19) in Theorem 3.4.1, the behavior of the last factor in expression (4.41), $\sqrt{n}\,\mathcal{W}^{-1/2}\widehat{\mathbf{G}}_M$, is described by

$$(\mathbf{I}_{Mm^2} - \mathbf{P}_M)\sqrt{n}\,\mathcal{W}^{-1/2}\mathbf{G}_M + O_P(\frac{1}{\sqrt{n}}) \ ,$$

where $\mathbf{G}_M = \{[\mathrm{vec}(\mathbf{C}'_1)]', \ldots, [\mathrm{vec}(\mathbf{C}'_M)]'\}'$. Therefore, the component $A_n^M(u)$ is selected as the dominant term in (4.41). Hence,

$$A_n^M(u) = \boldsymbol{\alpha}'_{M-(p+q)}(u)[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m]\boldsymbol{\Psi}'_M\widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}(\mathbf{I}_{Mm^2} - \mathbf{P}_M)\sqrt{n}\,\mathcal{W}^{-1/2}\mathbf{G}_M \ . \tag{4.42}$$

With this choice for $A_n^M(u)$, the remainder process can be decomposed as

$$R_n^M(u) = P_n^M(u) + Q_n^M(u) \ , \quad 0 \le u \le 1 \ , \tag{4.43}$$

where

$$P_n^M(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=M+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}\frac{\sin(K\pi u)}{K} \ ;$$

and $Q_n^M(u)$ is a term bounded in probability.

**Second part.** Proceed as in Theorem 2.6.1, and select

$$A^M(u) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{M-(p+q)} v_k \frac{\sin(k\pi u)}{k} \quad , \quad R^M(u) = \frac{\sqrt{2}}{\pi}\sum_{k>M-(p+q)}^{\infty} v_k \frac{\sin(k\pi u)}{k} \ .$$

In this manner, $B(u) = A^M(u) + R^M(u)$.

**Third part.** The rest of this proof consists in checking that conditions **(C.1)**$-$**(C.2)**$-$**(C.3)** of Lemma 2.6.1 hold for $M \ge M_0 = p+q+1$.

**(C.1)** For $r \ge 1$ and $u_1, u_2, \ldots, u_r$ in $[0,1]$, from (4.42) it can be written

$$\begin{pmatrix} A_n^d(u_1) \\ A_n^d(u_2) \\ \vdots \\ A_n^M(u_r) \end{pmatrix} = \mathbf{M}[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m]\boldsymbol{\Psi}'_M \widehat{\mathbf{W}}^{-1/2}\mathcal{W}^{1/2}(\mathbf{I}_{Mm^2} - \mathbf{P}_M)\sqrt{n}\mathcal{W}^{-1/2}\mathbf{G}_M \ ,$$

where $\mathbf{M}$ is a $r \times [M-(p+q)]$ constant matrix whose $jth$ row is $\boldsymbol{\alpha}'_{M-(p+q)}(u_j)$, $j = 1, \ldots, r$. From parts (a)$-$(b) of Proposition 4.3.1, it follows that $[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m]\boldsymbol{\Psi}'_M\boldsymbol{\Psi}_M[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}_m] = \mathbf{I}_{M-(p+q)}$; and $\boldsymbol{\Psi}'_M\mathbf{P}_M = \mathbf{0}$. Consider a sequence $\{\mathbf{V}_k : k \ge 1\}$ of i.i.d. random vectors $N_{m^2}(\mathbf{0}, \mathbf{I}_{m^2})$. For fixed $M$, from Proposition 2.6.1 and Slutsky's theorem, as $n \to \infty$

$$\begin{pmatrix} A_n^d(u_1) \\ A_n^d(u_2) \\ \vdots \\ A_n^M(u_r) \end{pmatrix} \xrightarrow{D} \mathbf{M}[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m]\boldsymbol{\Psi}'_M \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_M \end{pmatrix} \stackrel{D}{=}$$

$$\stackrel{D}{=} \mathbf{M}N_{M-(p+q)}[\mathbf{0}, \mathbf{I}_{M-(p+q)}] \stackrel{D}{=} \mathbf{M}\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{M-(p+q)} \end{pmatrix} = \begin{pmatrix} A^M(u_1) \\ A^M(u_2) \\ \vdots \\ A^M(u_r) \end{pmatrix} \ . \ \blacksquare$$

**(C.2)** Fix $M \geq M_0 = p + q + 1$. For checking tightness of the distributions of $\{A_n^M(u) : 0 \leq u \leq 1\}$, we proceed as in part **(C.2)** of Theorem 3.4.1. ■

**(C.3)** Verification of this condition for $R^M(u)$ has been done previously in Theorem 2.6.1. Derivations for the term $P_n^M(u)$ in (4.43), that are somewhat cumbersome, are presented in Appendix 4.1. ■

## 4.5 Consequences

From Theorem 4.4.1, under a *VARMA(p,q)* specification the asymptotic distribution of any continuous functional $H[\widehat{Z}_n^m(u)]$ of the modified process of (1.30) is given by $H[B(u)]$. The proposal then is to assess the goodness-of-fit of a model of the form (1.1) with the rejection criterion

$$H[\widehat{Z}_n^m(u)] \geq H_\alpha[B(u)] \ , \tag{4.44}$$

where $H_\alpha[B(u)]$ is the $(1-\alpha)-$ quantile of the distribution of $H[B(u)]$. The empirical size of region (4.44) will be approximately $\alpha$. From the discussion in section 3.5, (4.44) is bound to have better power properties than the analogue region of expression (3.34),

$$H[\widehat{W}_n^m(u)] \geq H_\alpha[B(u)] \ ,$$

that is based on the original residual process of (1.28). Comparisons between the regions (4.44) and (3.34) will be analyzed in the simulation examples of chapter 5.

Goodness-of-fit functionals considered are the Kolmogorov-Smirnov criterion,

$$\sup_{0 \leq u \leq 1} |\widehat{Z}_n^m(u)| \ ; \tag{4.45}$$

and the Cramér-von Mises statistic,

$$CVM = \int_0^1 [\widehat{Z}_n^m(u)]^2 du = \frac{n}{m\,\pi^2} \sum_{k=p+q+1}^{n-(P+1)} \frac{[\mathrm{tr}(\widehat{\mathbf{S}}_k)]^2}{K^2} \ . \tag{4.46}$$

In practice, criterion (4.45) is approximated by

$$KS = \sup_{1 \leq j \leq n} |\widehat{Z}_n^m(j/n)| \ . \tag{4.47}$$

Using the tightness condition for the distributions of $\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$, the asymptotic distribution of $KS$ is that of (4.45), $\sup_{0 \leq u \leq 1} |B(u)|$. Another possibility is to approximate the criterion $CVM$ in (4.46) by the Riemann sum

$$PCVM = \frac{1}{n} \sum_{j=1}^{n} [\widehat{Z}_n^m(j/n)]^2 \ . \tag{4.48}$$

The limit distribution of both $CVM$ and $PCVM$ is $\int_0^1 [B(u)]^2 du$. The behavior of these goodness-of-fit criteria, $KS$ in (4.47); $CVM$ in (4.46); and $PCVM$ in (4.48), is analyzed in chapter 5.

## Appendix 4.1: Proof of condition (C.3) in Theorem 4.4.1

From definition (4.28),

$$\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m} = \mathbf{a}'_m \mathrm{vec}(\widehat{\mathbf{S}}_k) = \mathbf{a}'_m \widehat{\mathbf{\Gamma}}'_k \widehat{\mathbf{W}}^{-1/2} \widehat{\mathbf{G}}_k \ , \tag{4.49}$$

where, using expression (2.27), $\widehat{\mathbf{G}}_k = \mathbf{G}_k - \mathbf{Z}_k \mathrm{vec}[(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}}) - (\mathbf{\Phi}, \mathbf{\Theta})] + O_P(1/n)$. Thus, $P_n^M(u)$ in (4.43) can be decomposed in the form

$$P_n^M(u) = \frac{\sqrt{2}}{\pi} \sqrt{n} \sum_{k=M+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}} \frac{\sin(K\pi u)}{K} =$$

$$= C_n^M(u) + D_n^M(u) \ , \quad 0 \leq u \leq 1 \ , \tag{4.50}$$

where

$$C_n^M(u) = \frac{\sqrt{2}}{\pi} \sqrt{n} \sum_{k=M+1}^{n-(P+1)} \mathbf{a}'_m \widehat{\mathbf{\Gamma}}'_k \widehat{\mathbf{W}}^{-1/2} \mathbf{G}_k \frac{\sin(K\pi u)}{K} \ ,$$

and $D_n^M(u)$ is a bounded remainder term.

For checking **(C.3)** for $C_n^M(u)$ in (4.50), consider the decomposition

$$C_n^M(u) = X_n^M(u) + Y_n^M(u) + Z_n^M(u) \ , \quad 0 \leq u \leq 1 \ , \tag{4.51}$$

where

$$X_n^M(u) = \frac{\sqrt{2}}{\pi} \sqrt{n} \sum_{k=M+1}^{n-(P+1)} \mathbf{a}'_m \mathbf{E}'_k \widehat{\mathbf{W}}^{-1/2} \mathbf{G}_k \frac{\sin(K\pi u)}{K} \ ;$$

$$Y_n^M(u) = \frac{\sqrt{2}}{\pi} \sqrt{n} \sum_{k=M+1}^{n-(P+1)} \mathbf{a}'_m (\widehat{\mathbf{\Gamma}}_k - \mathbf{\Gamma}_k)' \widehat{\mathbf{W}}^{-1/2} \mathbf{G}_k \frac{\sin(K\pi u)}{K} \ ;$$

and

$$Z_n^M(u) = \frac{\sqrt{2}}{\pi} \sqrt{n} \sum_{k=M+1}^{n-(P+1)} \mathbf{a}'_m (\mathbf{\Gamma}_k - \mathbf{E}_k)' \widehat{\mathbf{W}}^{-1/2} \mathbf{G}_k \frac{\sin(K\pi u)}{K} \ ,$$

where

$$\mathbf{E}_k = \begin{pmatrix} \mathbf{0}_{(k-1)m^2 \times m^2} \\ \hline \mathbf{I}_{m^2} \end{pmatrix}$$

is the $km^2 \times m^2$ matrix of expression (4.31) in Lemma 4.3.1.

On one hand,

$$\mathbf{a}'_m \mathbf{E}'_k \widehat{\mathbf{W}}^{-1/2} \mathbf{G}_k = [\mathbf{0}_{1 \times (k-1)m^2} | \mathrm{vec}(\mathbf{I}_m)'] \widehat{\mathbf{W}}^{-1/2} \mathbf{G}_k / \sqrt{m} =$$

$$= \mathrm{vec}(\mathbf{I}_m)' \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2} \mathbf{C}'_k \widehat{\mathbf{\Sigma}}^{-1/2}) / \sqrt{m} = \mathrm{tr}(\mathbf{C}'_k \widehat{\mathbf{\Sigma}}^{-1}) / \sqrt{m} \ .$$

Thus, $X_n^M(u)$ in (4.51) can be processed as the term $R_n^M(u)$ in Theorem 2.6.1, and using the fact that $\widehat{\mathbf{\Sigma}}$ is consistent for $\mathbf{\Sigma}$.

For dealing with the terms $Y_n^M(u)$ and $Z_n^M(u)$ in expression (4.51), the following two results are needed.

**Lemma 4.5.1** Consider the sequence of matrices $\{\mathbf{\Gamma}_k : p+q+1 \leq k \leq n-(P+1)\}$ of expression (4.23). Then:

(a) There exists an open neighborhood $U \subset \mathbb{R}^{m^2(p+q)}$ of radius $\delta > 0$ centered at $\mathbf{\Lambda} = \mathrm{vec}(\mathbf{\Phi}, \mathbf{\Theta})$ such that, if $\widehat{\mathbf{\Lambda}} = \mathrm{vec}(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}}) \in U$,

$$\|\mathbf{a}_m'(\widehat{\mathbf{\Gamma}}_k - \mathbf{\Gamma}_k)'\| \leq ab^k \|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\| \ , \tag{4.52}$$

for $k = 1+p+q, \ldots, n-(P+1)$, where $a > 0$, and $0 < b < 1$.

(b) For $k = 1+p+q, \ldots, n-(P+1)$,

$$\|\mathbf{a}_m'(\mathbf{\Gamma}_k - \mathbf{E}_k)'\| \leq ab^k \ , \tag{4.53}$$

where the constants $a > 0$, and $0 < b < 1$ are as in (4.52).

**Proof.** Inequalities (4.52) in part (a) follow by combining a Taylor's series expansion of $\widehat{\mathbf{\Gamma}}$ around $\mathbf{\Gamma}$ with the bounds of Appendix 3.1. Details are very lengthy, and are omitted for conciseness. Part (b) can be obtained similarly. ■

**Lemma 4.5.2** Consider the $[n-(P+1)]m^2 \times 1$ random vector

$$\mathbf{G}_{n-(P+1)} = \{[\mathrm{vec}(\mathbf{C}_1')]', [\mathrm{vec}(\mathbf{C}_2')]', \ldots, [\mathrm{vec}(\mathbf{C}_{n-(P+1)}')]'\}' \ ,$$

and put $\mathcal{W} = \mathbf{I}_{n-(P+1)} \otimes \mathbf{\Sigma} \otimes \mathbf{\Sigma}$. Then:

$$\mathrm{E}[\mathcal{W}^{-1/2}\mathbf{G}_{n-(P+1)}\mathbf{G}_{n-(P+1)}'\mathcal{W}^{-1/2}] = \frac{1}{n^2}\,\mathrm{diag}(n-1, n-2, \ldots, P+1) \otimes \mathbf{I}_{m^2} \ . \tag{4.54}$$

As a consequence of (4.54), $\mathrm{E}[\|\mathcal{W}^{-1/2}\mathbf{G}_{n-(P+1)}\|^2] \leq m^2$.

**Proof.** As established in Appendix 3.4,

$$\mathrm{E}\{\mathrm{vec}(\mathbf{C}_J')[\mathrm{vec}(\mathbf{C}_K')]'\} = \delta_{JK}\left(\frac{n-J}{n^2}\right)(\mathbf{\Sigma} \otimes \mathbf{\Sigma}) \ .$$

Consequently,

$$\mathrm{E}[\mathcal{W}^{-1/2}\mathbf{G}_{n-(P+1)}\mathbf{G}_{n-(P+1)}'\mathcal{W}^{-1/2}] = \mathcal{W}^{-1/2}\mathrm{E}[\mathbf{G}_{n-(P+1)}\mathbf{G}_{n-(P+1)}']\mathcal{W}^{-1/2} =$$

$$= \frac{1}{n^2}\,\mathrm{diag}(n-1, n-2, \ldots, P+1) \otimes \mathbf{I}_{m^2} \ .$$

This establishes identity (4.54). From there,

$$\mathrm{E}[\|\mathcal{W}^{-1/2}\mathbf{G}_{n-(P+1)}\|^2] = \mathrm{tr}\{\mathrm{E}[\mathcal{W}^{-1/2}\mathbf{G}_{n-(P+1)}\mathbf{G}'_{n-(P+1)}\mathcal{W}^{-1/2}]\} =$$

$$= m^2[\frac{1}{n^2}\sum_{j=1}^{n-(P+1)}(n-j)] = \frac{m^2}{n^2}\{n[n-(P+1)] - \frac{(n-P)[n-(P+1)]}{2}\} \leq$$

$$\leq m^2\frac{n[n-(P+1)]}{n^2} \leq m^2 \ . \ \blacksquare$$

The treatment of the processes $Y_n^M(u)$ and $Z_n^M(u)$ in expression (4.51) can be done by first considering

$$\overline{Y}_n^M(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=M+1}^{n-(P+1)}\mathbf{a}'_m(\widehat{\mathbf{\Gamma}}_k - \mathbf{\Gamma}_k)'\mathcal{W}^{-1/2}\mathbf{G}_k\frac{\sin(K\pi u)}{K} \ ; \qquad (4.55)$$

and

$$\overline{Z}_n^d(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=M+1}^{n-(P+1)}\mathbf{a}'_m(\mathbf{\Gamma}_k - \mathbf{E}_k)'\mathcal{W}^{-1/2}\mathbf{G}_k\frac{\sin(K\pi u)}{K} \ ; \qquad (4.56)$$

and then using consistency of $\widehat{\mathbf{W}}$ to $\mathcal{W}$.

By applying Lemmas 4.5.1 and 4.5.2 and the Cauchy-Schwarz inequality, it follows that the process of (4.55) verifies:

$$\sup_{0\leq u\leq 1}|\overline{Y}_n^M(u)| \leq \frac{\sqrt{2}}{\pi}\sqrt{n}(\sum_{k=M+1}^{\infty}1/K^2)^{1/2}\{\sum_{k=M+1}^{n-(P+1)}[\mathbf{a}'_m(\widehat{\mathbf{\Gamma}}_k - \mathbf{\Gamma}_k)'\mathcal{W}^{-1/2}\mathbf{G}_k]^2\}^{1/2} \leq$$

$$\leq \frac{\sqrt{2}}{\pi}\sqrt{n}(\sum_{k=M+1}^{\infty}1/K^2)^{1/2}[\sum_{k=M+1}^{n-(P+1)}a^2b^{2k}\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\|^2\|\mathcal{W}^{-1/2}\mathbf{G}_{n-(P+1)}\|^2]^{1/2} \leq$$

$$\leq \frac{a\sqrt{2}}{\pi(1-b^2)}b^{M+1}(\sum_{k=M+1}^{\infty}1/K^2)^{1/2}L_n \ , \qquad (4.57)$$

where $L_n = \sqrt{n}\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\|\|\mathcal{W}^{-1/2}\mathbf{G}_{n-(P+1)}\| = O_P(1)$. In turn, for (5.54)

$$\sup_{0\leq u\leq 1}|\overline{Z}_n^M(u)| \leq \frac{\sqrt{2}}{\pi}(\sum_{k=M+1}^{\infty}1/K^2)^{1/2}(n\sum_{k=M+1}^{n-(P+1)}\boldsymbol{\eta}_k)^{1/2} \ , \qquad (4.58)$$

where $\boldsymbol{\eta}_k = [\mathbf{a}'_m(\mathbf{\Gamma}_k - \mathbf{E}_k)'\mathcal{W}^{-1/2}\mathbf{G}_k]^2$. From the proof of Lemma 4.5.2, $\mathcal{W}^{-1/2}\mathbf{G}_k$ is a random vector with mean $\mathbf{0}$ and block-diagonal covariance matrix with generic block

$[(n-j)/n^2]\mathbf{I}_{m^2}$, where $(n-j)/n^2 \leq n^{-1}$, $j = 1,\ldots, k$. As a consequence, it is easily obtained after some algebra that

$$n\mathrm{E}(\boldsymbol{\eta}_k) = n\mathrm{tr}[(\boldsymbol{\Gamma}_k - \mathbf{E}_k)\mathbf{a}_m\mathbf{a}'_m(\boldsymbol{\Gamma}_k - \mathbf{E}_k)'\mathrm{Var}(\mathcal{W}^{-1/2}\mathbf{G}_k)] =$$

$$= n\mathbf{a}'_m(\boldsymbol{\Gamma}_k - \mathbf{E}_k)'\mathrm{Var}(\mathcal{W}^{-1/2}\mathbf{G}_k)(\boldsymbol{\Gamma}_k - \mathbf{E}_k)\mathbf{a}_m \leq \|\mathbf{a}'_m(\boldsymbol{\Gamma}_k - \mathbf{E}_k)'\|^2 \leq a^2b^{2k} \ . \quad (4.59)$$

From inequalities (4.57)−(4.58)−(4.59), it easy to see that (**C.3**) is satisfied for both $\overline{Y}_n^M(u)$ and $\overline{Z}_n^M(u)$ in (4.55) and (5.54), respectively. ■

# Chapter 5

# Examples, simulations, and comparisons

**Summary.** This chapter studies the practical properties of the methods of chapters 2, 3, and 4. Section 5.1 gives a general overview of the techniques used. Section 5.2 contains an initial exploration for several *VARMA(p,q)* models. In section 5.3, the behavior of the adjusted traces is studied: the residual $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ of (1.29); the modified $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ of (4.11); and, for completeness, the model error version $\mathrm{tr}(\mathbf{R}_k)/\sqrt{m}$ of (2.43). Section 5.4 analyzes in more detail the empirical behavior, in both size and power, of the functionals $KS$ of (4.47); $CVM$ of (4.46); and $PCVM$ of (4.48). These are used with the different goodness-of-fit processes studied in this thesis: the error process $\{W_n^m(u) : 0 \leq u \leq 1\}$ of (2.44); the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (1.28); and the modified process $\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$ of (1.30). Section 5.5 compares the same functionals, when they are applied to the truncated version (4.33) of $\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$, with the standard criteria of Hosking (1980) in (1.22); and Li and McLeod (1981) in (1.24). Section 5.6 defines a multivariate version of the standard cumulative periodogram statistic. Section 5.7 discusses an application on a well-known trivariate quarterly time series on West German investment, income, and consumption between 1960−1982. Section 5.8 contains some final conclusions.

## 5.1 Introduction

In this chapter, the results of the previous parts of this thesis are illustrated. Simulation techniques are used to compare the different procedures. Numerical examples are implemented with specifically constructed Fortran 90 codes. This software adapts suitably to the complexity of the transformation methods of chapter 4. Other computer programs for multivariate time series appear in standard packages such as GAUSS, MATLAB, SAS, SPLUS, .... A recent collection of R routines for multivariate time series was given by Mahdi and McLeod (2013).

For the case of pure autoregressive *VAR(p)* models, the Yule-Walker estimators $\widehat{\mathbf{\Phi}}_i$, $i = 1, \ldots, p$, are used. These are obtained as solutions of the normal linear equations of (2.10). For *VARMA(p,q)* processes with $q > 0$, the scoring iteration of expression (2.17) of section 2.2.2,

$$\mathbf{\Lambda}^{k+1} = \mathbf{\Lambda}^k - s^k \mathbf{I}_n^{k^{-1}} [\frac{\partial l_n(\mathbf{\Lambda}, \overline{\mathbf{X}}_n)}{\partial \mathbf{\Lambda}}|_{\mathbf{\Lambda}=\mathbf{\Lambda}^k}] \ ,$$

is considered. The above algorithm is the ML estimation scheme suggested by Lütkepohl (2005, sec.12.3). Initial values $\mathbf{\Lambda}^0 = \text{vec}(\mathbf{\Phi}^0, \mathbf{\Theta}^0)$ and $\mathbf{\Sigma}_0$ are selected naturally as the true parameters of the population model used in the simulation. The step length $s^k$ is taken to be equal to a fixed constant $0 < s < 1$, independent of the particular iteration considered. The matrix $\mathbf{I}_n^k$ is sometimes found to be ill-conditioned. Therefore, before proceeding to its inversion, it is regularized by adding a small positive constant to its diagonal elements. In general, finding the ML estimates of the parameters of a *VARMA(p,q)* model is a complicated task, and there is not a universally accepted solution in practice. However, we have found in our simulations that (2.17) converges after an adequate number of iterations. Thus, it seems to be a suitable tool for illustrating our results of chapters 3 and 4.

Once the ML estimates $(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}})$ of the parameters $(\mathbf{\Phi}, \mathbf{\Theta})$ have been determined, the $m \times 1$ residuals $\widehat{\boldsymbol{\varepsilon}}_t$, $P < t \leq n$, are computed using the recursion of (2.4), where $P = \max(p, q)$. The ML estimate of the covariance matrix $\mathbf{\Sigma}$ is as given in (2.18),

$$\widehat{\mathbf{\Sigma}} = \frac{1}{n} \sum_{t>P}^{n} \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t' \ .$$

The $m \times m$ residual autocorrelation matrices of Chitturi (1974) are computed as defined in (1.21),

$$\widehat{\mathbf{R}}_k = \widehat{\mathbf{C}}_k' \widehat{\mathbf{\Sigma}}^{-1} \ , \quad 1 \leq k \leq n - (P+1) \ ,$$

where the $m \times m$ matrices

$$\widehat{\mathbf{C}}_k = \frac{1}{n} \sum_{t>P}^{n-k} \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_{t+k}' \ , \quad 0 \leq k \leq n - (P+1) \ ,$$

are as in (1.20). Finding the adjusted residual traces $\text{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ of (1.29), $1 \leq k \leq n - (P+1)$, is then relatively straightforward.

In turn, the modified adjusted residual traces of (4.11),

$$\text{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m} = \mathbf{a}_m' \text{vec}(\widehat{\mathbf{S}}_k) \ , \qquad p+q+1 \leq k \leq n - (P+1) \ ,$$

where $\mathbf{a}_m = \text{vec}(\mathbf{I_m})/\sqrt{m}$ is of $m^2 \times 1$, are computed using the recursion of (4.29),

$$\text{vec}(\widehat{\mathbf{S}}_k) = [\mathbf{I}_{m^2} + \widehat{\mathbf{\Xi}}_k' (\widehat{\mathbf{Z}}_{k-1}' \widehat{\mathcal{W}}^{-1} \widehat{\mathbf{Z}}_{k-1})^{-1} \widehat{\mathbf{\Xi}}_k]^{-1/2}$$

$$[\,\text{vec}(\widehat{\mathbf{\Sigma}}^{-1/2} \widehat{\mathbf{C}}_k' \widehat{\mathbf{\Sigma}}^{-1/2}) - \widehat{\mathbf{\Xi}}_k' (\widehat{\mathbf{Z}}_{k-1}' \widehat{\mathcal{W}}^{-1} \widehat{\mathbf{Z}}_{k-1})^{-1} \sum_{j=1}^{k-1} \widehat{\mathbf{\Xi}}_j \text{vec}(\widehat{\mathbf{\Sigma}}^{-1/2} \widehat{\mathbf{C}}_j' \widehat{\mathbf{\Sigma}}^{-1/2})\,] \ ,$$

$p+q+1 \leq k \leq n-(P+1)$. The estimated $m^2 \times m^2(p+q)$ row blocks $\widehat{\boldsymbol{\Xi}}'_k$ are determined by replacing unknown parameters by estimators in definition (2.23),

$$\boldsymbol{\Xi}'_k = (\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2})(\mathbf{G}_{k-1}, \ldots, \mathbf{G}_{k-p}\,; \mathbf{F}_{k-1}, \ldots, \mathbf{F}_{k-q}) \ , \quad k \geq 1 \ ,$$

where $\mathbf{G}_k = \sum_{j=0}^{k}(\boldsymbol{\Sigma}\boldsymbol{\Omega}'_j \otimes \mathbf{L}_{k-j})$, $\mathbf{F}_k = \boldsymbol{\Sigma} \otimes \mathbf{L}_k$, $k \geq 0$; and $\mathbf{G}_k = \mathbf{F}_k = \mathbf{0}$, $k < 0$. The $m \times m$ matrices $\{\boldsymbol{\Omega}_j : j \geq 0\}$ and $\{\mathbf{L}_j : j \geq 0\}$ are the coefficients of the series expansions $\boldsymbol{\Phi}^{-1}(z)\boldsymbol{\Theta}(z) = \sum_{j=0}^{\infty} \boldsymbol{\Omega}_j z^j$ and $\boldsymbol{\Theta}^{-1}(z) = \sum_{j=0}^{\infty} \mathbf{L}_j z^j$, where $\boldsymbol{\Omega}_0 = \mathbf{L}_0 = \mathbf{I}_m$. On the other hand, $\widehat{\mathbf{Z}}'_k \widehat{\mathcal{W}}^{-1} \widehat{\mathbf{Z}}_k = \sum_{j=1}^{k} \widehat{\boldsymbol{\Xi}}_k \widehat{\boldsymbol{\Xi}}'_k$.

The $\boldsymbol{\Omega}_j$ and $\mathbf{L}_j$ can be determined, using finite recursive schemes with suitable initial conditions, from the initial matrix parameters $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_p)$ and $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \cdots, \boldsymbol{\Theta}_q)$ of model (1.1). For the coefficients $\boldsymbol{\Omega}_j$, it easily follows that:

$$\begin{aligned} \boldsymbol{\Theta}_j \ &= \ \boldsymbol{\Omega}_j - \boldsymbol{\Phi}_1 \boldsymbol{\Omega}_{j-1} - \cdots - \boldsymbol{\Phi}_p \boldsymbol{\Omega}_{j-p} \ , \quad 1 \leq j \leq q \ ; \\ \mathbf{0}_{m\times m} \ &= \ \boldsymbol{\Omega}_j - \boldsymbol{\Phi}_1 \boldsymbol{\Omega}_{j-1} - \cdots - \boldsymbol{\Phi}_p \boldsymbol{\Omega}_{j-p} \ , \quad j > q \ ; \end{aligned} \tag{5.1}$$

The recursion of (5.1) can be solved with the aid of the conditions $\boldsymbol{\Omega}_0 = \mathbf{I}_m$, and $\boldsymbol{\Omega}_j = \mathbf{0}_{m\times m}$, $j < 0$. In particular, $\boldsymbol{\Omega}_j = \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{\Omega}_{j-i}$, $j > \max(p,q) = P$. The matrices $\mathbf{L}_j$ can be determined similarly.

In general, we have found the numerical behavior of (4.29) quite tractable. In some particular models, the matrix $\widehat{\mathbf{Z}}'_{k-1} \widehat{\mathcal{W}}^{-1} \widehat{\mathbf{Z}}_{k-1}$ is ill-conditioned for the initial values of $k \geq p+q+1$. When this occurs, the inverse $(\widehat{\mathbf{Z}}'_{k-1} \widehat{\mathcal{W}}^{-1} \widehat{\mathbf{Z}}_{k-1})^{-1}$ is taken after adding a small positive constant to the diagonal elements.

The method used for simulating $m\times 1$ data vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from a *VARMA(p,q)* model $\boldsymbol{\Phi}(B)(\mathbf{X}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(B)\boldsymbol{\varepsilon}_t$ of the form (1.1), with known parameters $\boldsymbol{\mu}$ and $(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma})$, is as follows:

**(a)** A collection of $2n$ i.i.d. random vectors $\boldsymbol{\varepsilon}_t$ with distribution $N_m(\mathbf{0}, \boldsymbol{\Sigma})$ is generated.

**(b)** An auxiliary sequence of $m \times 1$ data vectors $\mathbf{Y}_1, \ldots, \mathbf{Y}_{2n}$ is constructed using the equations

$$\boldsymbol{\Phi}(B)(\mathbf{Y}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(B)\boldsymbol{\varepsilon}_t \ ;$$

and the initial conditions $\mathbf{Y}_t - \boldsymbol{\mu} \equiv \mathbf{0} \equiv \boldsymbol{\varepsilon}_t$, $t \leq 0$.

**(c)** The first $n$ $\mathbf{Y}_t$ are removed, and $\mathbf{X}_t = \mathbf{Y}_{t+n}$, $1 \leq t \leq n$.

This is a standard method that eliminates the dependence of the observations $\mathbf{X}_t$ from the initial conditions $\mathbf{Y}_t - \boldsymbol{\mu} \equiv \mathbf{0} \equiv \boldsymbol{\varepsilon}_t$, $t \leq 0$. In the simulations presented below, the mean vector $\boldsymbol{\mu} = \mathrm{E}(\mathbf{X}_t)$ is always taken to be equal to $\mathbf{0}$.

The statistics $KS$ of (4.47); $CVM$ of (4.46); and $PCVM$ of (4.48) are compared to the adequate critical points of the distributions of $\sup_{0 \leq u \leq 1} |B(u)|$ and $\int_0^1 [B(u)]^2 du$, respectively. For a nominal significance level $0 < \alpha < 1$, the notation for the corresponding $(1-\alpha) \times 100\%$ quantiles will be $KS_\alpha$ and $CVM_\alpha$, respectively. Nominal levels $\alpha = .1$, $.05$, and $.01$ will be used. From standard tables (Shorack and Wellner, 1986), it is obtained that $KS, .1 = 1.2238$; $KS, .05 = 1.3582$; and $KS, .01 = 1.6277$. Also, $CVM, .1 = 0.3473$; $CVM, .05 = 0.4614$; and $CVM, .01 = 0.7435$.

## 5.2 Examples of *VARMA(p,q)* processes

Mahdi and McLeod (2012, section 3) study several specifications of *VARMA(p,q)* processes analyzed earlier in the literature. These include the bivariate *VAR(1)* models considered by Hosking (1980), and Li and McLeod (1981); the $m = 2$ *VARMA(1,1)* models of Brockwell and Davis (1991, p. 428), and Reinsel (1997, p. 81); the two dimensional *VMA(1)* of Reinsel (1997, p. 25); a *VAR(2)* with $m = 2$ in Lütkepohl (2005, p.17), and a bivariate *VARMA(2,1)* by Lütkepohl (2005, p.445).

In this section, we use a collection of new *VARMA(p,q)* models to illustrate the theoretical and empirical properties of the techniques of chapters 2, 3, and 4.

### 5.2.1 The *VAR(1)* model

Consider the *VAR(1)* model for $m = 2$,

$$\mathbf{X}_t = \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t \ , \tag{5.2}$$

where

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.2802 & 0.2680 \\ -0.0183 & 0.3152 \end{pmatrix} . \tag{5.3}$$

The matrix of (5.3) is obtained by taking eigenvalues $\delta_j = 0.2977 \pm 0.0678\, i$, $j = 1, 2$, so that $|\delta_1| = |\delta_2| = 0.3053 < 1$. The associated eigenvectors are selected in the form $\boldsymbol{\gamma}_1 = (2.7071, 0.1768 + 0.6847\, i)'$, and $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (2.7071, 0.1768 - 0.6847\, i)'$. Thus, the array of (5.3) follows from the identity $\mathbf{\Phi}_1 = \mathbf{C}\mathbf{D}\mathbf{C}^{-1}$, where $\mathbf{C} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ and $\mathbf{D} = \mathrm{diag}(\delta_1, \delta_2)$. The covariance matrix of the errors $\boldsymbol{\varepsilon}_t$ in (5.2) will be given by

$$\mathbf{\Sigma} = \begin{pmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{pmatrix} . \tag{5.4}$$

From the discussion in section 4.3.4, the $j$th row-block of $\mathcal{W}^{-1/2}\mathbf{Z}_M$ is of the form

$$\mathbf{\Xi}_j' = \mathbf{\Sigma}^{1/2}(\mathbf{\Phi}_1^{j-1})' \otimes \mathbf{\Sigma}^{-1/2} \ , \quad j \geq 1 \ .$$

For $M$ large enough, the information matrix can be approximated by the sum

$$\mathbf{Z}_M'\mathcal{W}^{-1}\mathbf{Z}_M = \sum_{j=1}^{M} \mathbf{\Xi}_j \mathbf{\Xi}_j' = [\sum_{j=0}^{M-1} \mathbf{\Phi}_1^j \mathbf{\Sigma}(\mathbf{\Phi}_1^j)'] \otimes \mathbf{\Sigma}^{-1} \ ,$$

given in (4.37), where by convention $\mathbf{\Phi}_1^0 = \mathbf{I}_m$. It is found numerically that $\|\mathbf{\Xi}_j\| < 10^{-15}$ for $j > 6$. Thus, only $M = 6$ terms are really needed in the sum above. This produces the following information matrix:

$$\mathbf{I}[\mathrm{vec}(\mathbf{\Phi}_1)] = \begin{pmatrix} 1.3628 & -0.4088 & 0.4640 & -0.1392 \\ -0.4088 & 1.3628 & -0.1392 & 0.4640 \\ 0.4640 & -0.1392 & 1.2147 & -0.3644 \\ -0.1392 & 0.4640 & -0.3644 & 1.2147 \end{pmatrix} . \tag{5.5}$$

The structure of (5.5) resembles to that of the matrix

$$\mathbf{\Sigma} \otimes \mathbf{\Sigma}^{-1} = \begin{pmatrix} 1.0989 & -0.3297 & 0.3297 & -0.0989 \\ -0.3297 & 1.0989 & -0.0989 & 0.3297 \\ 0.3297 & -0.0989 & 1.0989 & -0.3297 \\ -0.0989 & 0.3297 & -0.3297 & 1.0989 \end{pmatrix} , \tag{5.6}$$

where

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} 1.0989 & -0.3297 \\ -0.3297 & 1.0989 \end{pmatrix} . \tag{5.7}$$

In fact, (5.6) is just the first dominant term in the sum of the form (4.37).

It is of interest now to analyze the magnitude of the coefficients $\mathbf{a}_2'\mathbf{P}_{jk}\mathbf{a}_2$ that appear in the second summand of the covariance function of (3.13),

$$\gamma^2(u,v) = [\min(u,v) - uv] - \frac{2}{\pi^2} \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} (\mathbf{a}_2'\mathbf{P}_{jk}\mathbf{a}_2) \frac{\sin(j\pi u)}{j} \frac{\sin(k\pi v)}{k} \ ,$$

$0 \leq u,v \leq 1$, where $\mathbf{P}_{jk} = \mathbf{\Xi}_j'\mathbf{I}^{-1}[\mathrm{vec}(\mathbf{\Phi}_1)]\mathbf{\Xi}_k$, $j,k \geq 1$, and $\mathbf{a}_2 = \mathrm{vec}(\mathbf{I}_2)/\sqrt{2} = (1,0,0,1)'/\sqrt{2}$. As it turns out, these coefficients are negligible for $j,k > 6$. For $j,k \leq 6$, they are displayed in the $6 \times 6$ matrix below:

$$\begin{pmatrix} 0.8772 & 0.2523 & 0.0684 & 0.0172 & 0.0039 & 0.0000 \\ 0.2523 & 0.1036 & 0.0382 & 0.0131 & 0.0042 & 0.0000 \\ 0.0684 & 0.0382 & 0.0163 & 0.0062 & 0.0022 & 0.0000 \\ 0.0172 & 0.0131 & 0.0062 & 0.0025 & 0.0009 & 0.0000 \\ 0.0039 & 0.0042 & 0.0022 & 0.0009 & 0.0003 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix} . \tag{5.8}$$

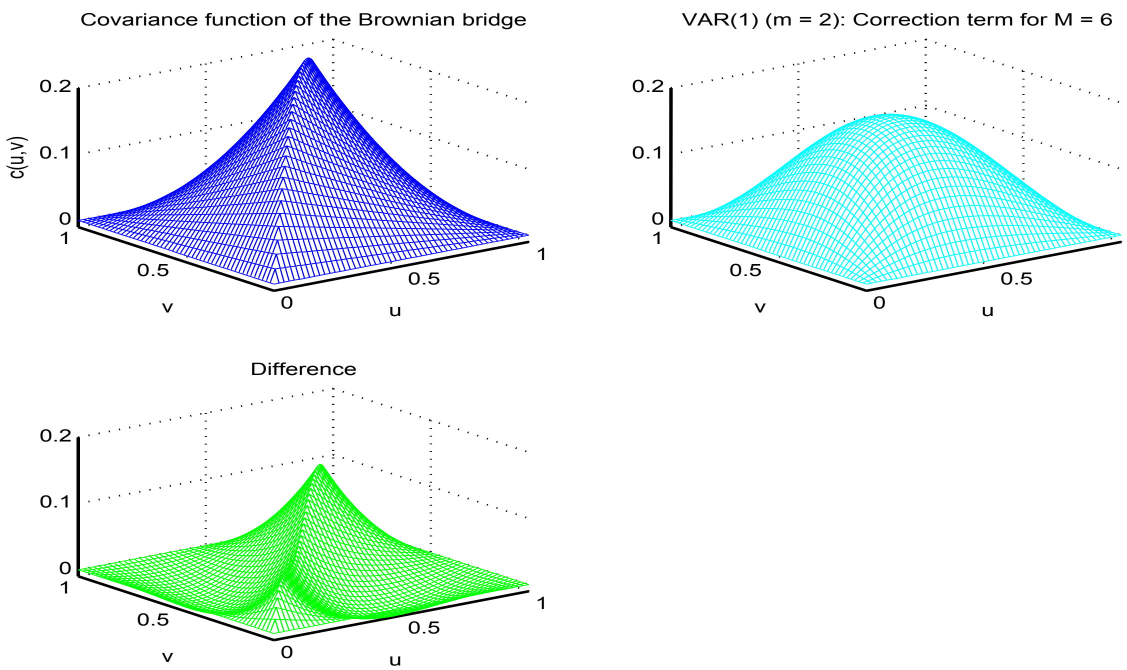


Figure 5.1: Covariance function for the limit of the residual process of (1.28) under the bivariate *VAR(1)* model (5.2) of section 5.2.1

Thus, in agreement with the discussion at the end of section 3.2.2, the covariance of the limit of the residual process $\{\widehat{W}_n^2(u) : 0 \leq u \leq 1\}$ of (1.28) behaves in the form

$$\gamma^2(u,v) = [\min(u,v) - uv] - \frac{2}{\pi^2} \sum_{j=1}^{M} \sum_{k=1}^{M} (\mathbf{a}_2' \mathbf{P}_{jk} \mathbf{a}_2) \frac{\sin(j\pi u)}{j} \frac{\sin(k\pi v)}{k} \ , \qquad (5.9)$$

where $M = 6$. Figure 5.1 displays all the functions that appear in (5.9). As seen there, $\gamma^2(u,v)$ is much smaller than the covariance function of the Brownian bridge, because of the substantial correction provided by the second summand at the right-hand side of (5.9). From the discussion of section 3.5, the rejection criteria $H[\widehat{W}_n^2(u)] \geq H_\alpha[B(u)]$ of expression (3.34) will have a size below the nominal level $\alpha$, and thus very low power. This will be analyzed in the simulation experiments of section 5.4.

From section 2.5, the asymptotic variances of the statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, where $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$ is the adjusted residual trace of (1.29), are equal to $1-$ the $k$th diagonal element of the matrix in (5.8), $k = 1, \ldots, 6$. These appear in the table below:

According to table 5.1, only the first asymptotic variance is really below 1. This corresponds to the leading diagonal element at the upper left corner of the matrix of (5.8), that takes the value 0.8772. A possible explanation for the pattern of table 5.1

| lag | var |
|---|---|
| 1 | 0.1228 |
| 2 | 0.8964 |
| 3 | 0.9837 |
| 4 | 0.9975 |
| 5 | 0.9997 |
| 6 | 1.0000 |

Table 5.1: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 6$, in the bivariate *VAR(1)* model (5.2) of section 5.2.1

is provided by the structure of the sum (4.37). It can be checked numerically that

$$\mathbf{P}_{11} = \begin{pmatrix} 0.8233 & 0.0000 & -0.0590 & 0.0000 \\ 0.0000 & 0.8233 & 0.0000 & -0.0590 \\ -0.0590 & 0.0000 & 0.9311 & 0.0000 \\ 0.0000 & -0.0590 & 0.0000 & 0.9311 \end{pmatrix} ; \tag{5.10}$$

and

$$\mathbf{P}_{22} = \begin{pmatrix} 0.1107 & 0.0000 & 0.0839 & 0.0000 \\ 0.0000 & 0.1107 & 0.0000 & 0.0839 \\ 0.0839 & 0.0000 & 0.1235 & 0.0000 \\ 0.0000 & 0.0839 & 0.0000 & 0.1235 \end{pmatrix} . \tag{5.11}$$

Hence, $\mathbf{a}_2'\mathbf{P}_{11}\mathbf{a}_2 = 0.8772$, and $\mathbf{a}_2'\mathbf{P}_{22}\mathbf{a}_2 = 0.1036$. The structure of both $\mathbf{P}_{11}$ and $\mathbf{P}_{22}$ can be explained by taking into account that from (4.37) and (5.5):

$$\mathbf{I}^{-1}[\mathrm{vec}(\mathbf{\Phi}_1)] \cong [\sum_{J=0}^{5} \mathbf{\Phi}_1^J \mathbf{\Sigma} (\mathbf{\Phi}_1^J)']^{-1} \otimes \mathbf{\Sigma} \ . \tag{5.12}$$

Also, $\mathbf{\Xi}_j' = \mathbf{\Sigma}^{1/2}(\mathbf{\Phi}_1^{j-1})' \otimes \mathbf{\Sigma}^{-1/2}$, $j \geq 1$. Hence, using (5.12)

$$\mathbf{P}_{jj} = \mathbf{\Xi}_j'\mathbf{I}^{-1}[\mathrm{vec}(\mathbf{\Phi}_1)]\mathbf{\Xi}_j = \mathbf{A}_j \otimes \mathbf{I}_2 \ , \tag{5.13}$$

where $\mathbf{A}_j \cong \mathbf{\Sigma}^{1/2}(\mathbf{\Phi}_1^{j-1})'[\sum_{J=0}^{5} \mathbf{\Phi}_1^J \mathbf{\Sigma} (\mathbf{\Phi}_1^J)']^{-1} \mathbf{\Phi}_1^{j-1} \mathbf{\Sigma}^{1/2}$, $j \geq 1$. Expression (5.13) explains the pattern of both the matrices $\mathbf{P}_{11}$ and $\mathbf{P}_{22}$ in (5.10) and (5.11), respectively. In particular, it can be checked that

$$\mathbf{A}_1 = \begin{pmatrix} 0.8233 & -0.0590 \\ -0.0590 & 0.9311 \end{pmatrix} \cong \mathbf{I}_2 \quad , \quad \mathbf{A}_2 = \begin{pmatrix} 0.1107 & 0.0839 \\ 0.0839 & 0.1235 \end{pmatrix} .$$

The considerations above can be extended for simulating *VAR(1)* models in dimension $m > 2$. For example, suppose starting eigenvalues $\delta_1 = 0.8500 + 0.1936\, i$; $\delta_2 = 0.8500 - 0.1936\, i$; and $\delta_3 = 0.4359$. The corresponding eigenvectors are taken as

$\boldsymbol{\gamma}_1 = (2.7071, 0.1768{+}0.6847\,i, 1.0000)'$; $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (2.7071, 0.1768{-}0.6847\,i, 1.0000)'$; and $\boldsymbol{\gamma}_3 = (1.0000, 1.0000, 1.0000)'$. Putting $\mathbf{C} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)$; $\mathbf{D} = \mathrm{diag}(\delta_1, \delta_2, \delta_3)$; and rescaling the expression $\mathbf{CDC}^{-1}$ by dividing by twice its Euclidean norm, leads to the $3 \times 3$ matrix

$$\boldsymbol{\Phi}_1 = \begin{pmatrix} 0.2673 & 0.1400 & -0.3275 \\ 0.0346 & 0.1646 & -0.1194 \\ 0.0693 & 0.0517 & -0.0413 \end{pmatrix} . \tag{5.14}$$

This can be used together with the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.0 & 0.3 & 0.3 \\ 0.3 & 1.0 & 0.3 \\ 0.3 & 0.3 & 1.0 \end{pmatrix} , \tag{5.15}$$

to form a trivariate *VAR(1)* model $\mathbf{X}_t = \boldsymbol{\Phi}_1 \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$, similar to (5.2).

Analogue comments to the ones given before for the structure of the parameter space of (5.2) apply. For instance, the asymptotic variances of the rescaled adjusted residual traces $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{3}$, $k = 1, \ldots, 5$, are displayed in the table below:

| lag | var |
|---|---|
| 1 | 0.0495 |
| 2 | 0.9533 |
| 3 | 0.9973 |
| 4 | 0.9999 |
| 5 | 1.0000 |

Table 5.2: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{3}$, $k = 1, \ldots, 5$, in the trivariate *VAR(1)* model (5.14)−(5.15) of section 5.2.1

### 5.2.2 Higher order vector autoregressive models

The construction of autoregressive *VAR(p)* models when $p > 1$ requires a criterion for relating the $m \times m$ matrices $\boldsymbol{\Phi}_i$, $i = 1, \ldots, p$, to a collection of prespecified roots, so that they are the solutions of the determinantal equation $|\boldsymbol{\Phi}(z)| = 0$, where $\boldsymbol{\Phi}(z) = \mathbf{I}_m - \boldsymbol{\Phi}_1 z - \cdots - \boldsymbol{\Phi}_p z^p$. The way to proceed is as follows:

**(a)** For each $j = 1, \ldots, m$, select roots $\varsigma_{j,i}$ with $|\varsigma_{j,i}| > 1$, $i = 1, \ldots, p$.

**(b)** For each $j = 1, \ldots, m$, form the polynomial of degree $p$:

$$p_j(z) = 1 - d_{j,1} z - d_{j,2} z^2 - \cdots - d_{j,p} z^p , \tag{5.16}$$

so that its roots are $\varsigma_{j,i}$, $i = 1, \ldots, p$.

**(c)** Construct the $m \times m$ diagonal matrices

$$\mathbf{D}_i = \mathrm{diag}(d_{1,i}, d_{2,i}, \ldots, d_{m,i}) \ , \quad i = 1, \ldots, p \ . \tag{5.17}$$

Recall that $\mathbf{D}_i$ is associated to the coefficients of the power $z^i$ in the polynomials $p_j(z)$ of (5.16), $j = 1, \ldots, m$.

**(d)** Consider an invertible matrix $\mathbf{A}$ of $m \times m$, and define

$$\mathbf{\Phi}_i = \mathbf{A}\mathbf{D}_i\mathbf{A}^{-1} \ , \quad i = 1, \ldots, p \ . \tag{5.18}$$

Under the construction (5.16)−(5.17)−(5.18), it follows that

$$|\mathbf{\Phi}(z)| = |\mathbf{I}_m - \mathbf{\Phi}_1 z - \cdots - \mathbf{\Phi}_p z^p| = |\mathbf{I}_m - \mathbf{D}_1 z - \cdots - \mathbf{D}_p z^p| = \prod_{j=1}^{m} p_j(z) \ . \tag{5.19}$$

By (5.19), the $mp$ roots of the determinantal equation $|\mathbf{\Phi}(z)| = 0$ are $\varsigma_{j,i}$, $j = 1, \ldots, m$; $i = 1, \ldots, p$. These correspond to those of the polynomials $p_j(z)$ of (5.16).

| $j$ | $\varsigma_{j,1}$ | $\lvert\varsigma_{j,1}\rvert$ | $\varsigma_{j,2}$ | $\lvert\varsigma_{j,2}\rvert$ |
|---|---|---|---|---|
| 1 | $4.8989 + 4.8989\,i$ | 6.9281 | $4.8989 - 4.8989\,i$ | 6.9281 |
| 2 | $7.4282 + 0.0000\,i$ | 7.4282 | $8.9138 + 0.0000\,i$ | 8.9138 |
| 3 | $7.9282 + 0.0000\,i$ | 7.9282 | $9.5138 + 0.0000\,i$ | 9.5138 |

Table 5.3: Roots of the determinantal equation $|\mathbf{\Phi}(z)| = 0$ of the trivariate *VAR(2)* model (5.20)−(5.15) of section 5.2.2

As an application of the above algorithm, consider the construction of a trivariate *VAR(2)* model with roots as given in table 5.3. The coefficients $\{d_{j,i}\}$ of the two degree polynomials $p_j(z) = 1 - d_{j,1}z - d_{j,2}z^2$ of (5.16) are obtained from the identities:

$$\begin{aligned} d_{j,1} &= \frac{1}{\varsigma_{j,1}} + \frac{1}{\varsigma_{j,2}} \ ; \\ d_{j,2} &= -\frac{1}{\varsigma_{j,1}\varsigma_{j,2}} \ , \end{aligned}$$

$j = 1, 2, 3$. The invertible matrix $\mathbf{A}$ of step **(d)** above is selected in the form

$$\mathbf{A} = \begin{pmatrix} 1.2 & 0.4 & 0.3 \\ 0.3 & 1.0 & 0.3 \\ 0.3 & 0.3 & 1.0 \end{pmatrix} \ .$$

This leads to the $3 \times 3$ matrices:

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.1985 & 0.0180 & 0.0044 \\ -0.0113 & 0.2522 & -0.0029 \\ -0.0089 & 0.0082 & 0.2315 \end{pmatrix} \ , \quad \mathbf{\Phi}_2 = \begin{pmatrix} -0.0218 & 0.0021 & 0.0019 \\ -0.0018 & -0.0147 & 0.0010 \\ -0.0021 & 0.0001 & -0.0127 \end{pmatrix} \ . \tag{5.20}$$

The covariance matrix $\mathbf{\Sigma}$ is taken as in (5.15).

An important difference with the *VAR(1)* case appears. The leading coefficients $\mathbf{a}_3'\mathbf{P}_{jk}\mathbf{a}_3$ of the second summand of the covariance function of (3.13) are displayed in the $5 \times 5$ matrix below:

$$\begin{pmatrix} 0.9997 & 0.0036 & -0.0156 & -0.0036 & 0.0000 \\ 0.0036 & 0.9492 & 0.2156 & 0.0337 & 0.0000 \\ -0.0156 & 0.2156 & 0.0496 & 0.0014 & 0.0000 \\ -0.0036 & 0.0337 & 0.0014 & -0.0002 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix} . \tag{5.21}$$

It can also be checked numerically that

$$\mathbf{P}_{11} = \begin{pmatrix} 0.9996 & 0.0000 & 0.0000 \\ 0.0000 & 0.9998 & 0.0000 \\ 0.0000 & 0.0000 & 0.9998 \end{pmatrix} \otimes \mathbf{I}_3 \cong \mathbf{I}_9 \ , \tag{5.22}$$

and

$$\mathbf{P}_{22} = \begin{pmatrix} 0.9591 & -0.0018 & 0.0000 \\ -0.0018 & 0.9392 & -0.0003 \\ 0.0000 & -0.0003 & 0.9471 \end{pmatrix} \otimes \mathbf{I}_3 \cong \mathbf{I}_9 \ . \tag{5.23}$$

Accordingly, considering the unit vector $\mathbf{a}_3 = \text{vec}(\mathbf{I}_3)/\sqrt{3} = (1,0,0,0,1,0,0,0,1)'\sqrt{3}$, it is obtained that both $\mathbf{a}_3'\mathbf{P}_{11}\mathbf{a}_3$ and $\mathbf{a}_3'\mathbf{P}_{22}\mathbf{a}_3$ are close to 1. As seen in figure 5.2, the correction of the covariance function of the Brownian bridge for $M = 5$ is much stronger than the one observed in figure 5.1.

| $j$ | $\varsigma_{j,1}$ | $\lvert\varsigma_{j,1}\rvert$ | $\varsigma_{j,2}$ | $\lvert\varsigma_{j,2}\rvert$ | $\varsigma_{j,3}$ | $\lvert\varsigma_{j,3}\rvert$ |
|---|---|---|---|---|---|---|
| 1 | $4.0000 + 4.0000\,i$ | 5.6568 | $4.0000 - 4.0000\,i$ | 5.6568 | $6.2225 + 0.0000\,i$ | 6.2225 |
| 2 | $7.3882 + 0.0000\,i$ | 7.3882 | $8.6196 + 0.0000\,i$ | 8.6196 | $9.8510 + 0.0000\,i$ | 9.8510 |

Table 5.4: Roots of the determinantal equation $|\mathbf{\Phi}(z)| = 0$ of the bivariate *VAR(3)* model (5.24)−(5.4) of section 5.2.2

Another example appears in the construction of a bivariate *VAR(3)* model associated to the roots in table 5.4. The coefficients $\{d_{j,i}\}$ of the three degree polynomials $p_j(z) = 1 - d_{j,1}z - d_{j,2}z^2 - d_{j,3}z^3$ of (5.16) are obtained now from the identities:

$$\begin{aligned} d_{j,1} &= \frac{1}{\varsigma_{j,1}} + \frac{1}{\varsigma_{j,2}} + \frac{1}{\varsigma_{j,3}} \ ; \\ d_{j,2} &= -(\frac{1}{\varsigma_{j,1}\varsigma_{j,2}} + \frac{1}{\varsigma_{j,1}\varsigma_{j,3}} + \frac{1}{\varsigma_{j,2}\varsigma_{j,3}}) \ ; \\ d_{j,3} &= \frac{1}{\varsigma_{j,1}\varsigma_{j,2}\varsigma_{j,3}} \ , \end{aligned}$$

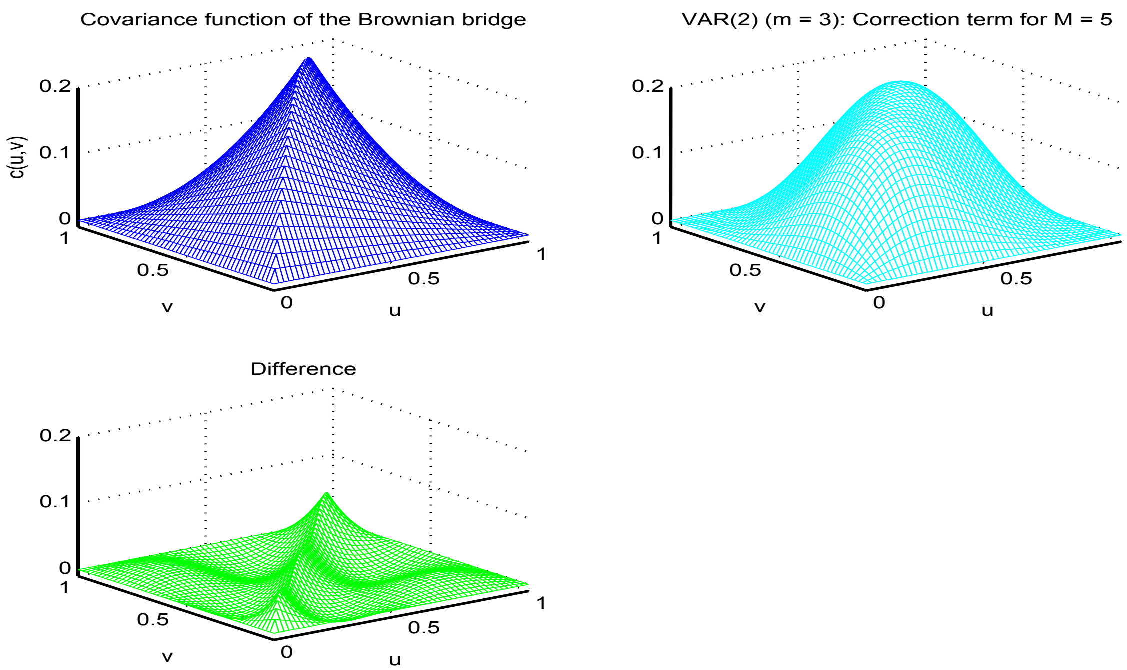


Figure 5.2: Covariance function for the limit of the residual process of (1.28) under the trivariate *VAR(2)* model (5.20)−(5.15) of section 5.2.2

$j = 1, 2$. The invertible matrix $\mathbf{A}$ of step **(d)** above is selected in the form

$$\mathbf{A} = \begin{pmatrix} 1.2 & 0.4 \\ 0.3 & 1.0 \end{pmatrix} .$$

This produces the $2 \times 2$ matrices

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.4171 & -0.0257 \\ 0.0161 & 0.3465 \end{pmatrix} \quad ; \quad \mathbf{\Phi}_2 = \begin{pmatrix} -0.0748 & 0.0134 \\ -0.0084 & -0.0379 \end{pmatrix} ;$$

$$\mathbf{\Phi}_3 = \begin{pmatrix} 0.0054 & -0.0015 \\ 0.0010 & 0.0012 \end{pmatrix} . \tag{5.24}$$

The covariance matrix $\mathbf{\Sigma}$ is as in (5.4). It can be checked that:

$$\begin{aligned} \mathbf{P}_{11} &= \begin{pmatrix} 0.9972 & 0.0009 \\ 0.0009 & 0.9939 \end{pmatrix} \otimes \mathbf{I}_2 ; \\ \mathbf{P}_{22} &= \begin{pmatrix} 0.9935 & 0.0077 \\ 0.0077 & 0.9808 \end{pmatrix} \otimes \mathbf{I}_2 ; \\ \mathbf{P}_{33} &= \begin{pmatrix} 0.8333 & -0.0051 \\ -0.0051 & 0.8821 \end{pmatrix} \otimes \mathbf{I}_2 , \end{aligned}$$

so that $\mathbf{P}_{jj} \cong \mathbf{I}_4$, $j = 1, 2, 3$. Thus, considering the vector $\mathbf{a}_2 = \mathrm{vec}(\mathbf{I}_2)/\sqrt{2} = (1, 0, 0, 1)'/\sqrt{2}$, for the bivariate *VAR(3)* model (5.24)−(5.4) there are now three leading coefficients $\mathbf{a}_2'\mathbf{P}_{jj}\mathbf{a}_2$ close to 1, $j = 1, 2, 3$.

A possible explanation for this structure of the first matrices $\mathbf{P}_{jj}$, $j = 1, \ldots, p$, when constructing *VAR(p)* models is given by the elementary matrix result below.

**Lemma 5.2.1** Consider a full-rank $m^2p$ matrix of the form $\mathbf{C} = (\mathbf{C}_1|\cdots|\mathbf{C}_p)$, where each $\mathbf{C}_i$ is of order $m^2p \times m^2$, $i = 1, \ldots, p$. Then:

$$\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}' = \mathrm{diag}(\mathbf{I}_{m^2}, \overset{(p)}{\ldots}, \mathbf{I}_{m^2}) \ . \tag{5.25}$$

As an application of expression (5.25) to a *VAR(p)* model, choose

$$\mathbf{C} = (\mathbf{C}_1|\cdots|\mathbf{C}_p) =$$

$$= \mathcal{W}^{-1/2}\mathbf{Z}_p = \mathcal{W}^{-1/2} \begin{pmatrix} \mathbf{G}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{G}_1 & \mathbf{G}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{G}_2 & \mathbf{G}_1 & \mathbf{G}_0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{p-1} & \mathbf{G}_{p-2} & \mathbf{G}_{p-3} & \cdots & \mathbf{G}_0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Xi}_1' \\ \vdots \\ \boldsymbol{\Xi}_p' \end{pmatrix} ,$$

where $\mathbf{G}_k = (\boldsymbol{\Sigma}\mathbf{H}_k' \otimes \mathbf{I}_m)$, $k \geq 0$; and the $\{\mathbf{H}_k : k \geq 0\}$ are the coefficients of the series expansion $\boldsymbol{\Phi}^{-1}(z) = \sum_{k=0}^{\infty} \mathbf{H}_k \mathbf{z}^k$. Thus, to some approximation it follows that

$$\mathcal{W}^{-1/2}\mathbf{Z}_p\mathbf{I}^{-1}[\mathrm{vec}(\boldsymbol{\Phi})]\mathbf{Z}_p'\mathcal{W}^{-1/2} \cong$$

$$\cong \mathcal{W}^{-1/2}\mathbf{Z}_p(\mathbf{Z}_p'\mathcal{W}^{-1}\mathbf{Z}_p)^{-1}\mathbf{Z}_p'\mathcal{W}^{-1/2} = \mathrm{diag}(\mathbf{I}_{m^2}, \overset{(p)}{\ldots}, \mathbf{I}_{m^2}) \ . \tag{5.26}$$

From (5.26) it is obtained roughly that for a *VAR(p)* model $\mathbf{P}_{jj} = \boldsymbol{\Xi}_j'\mathbf{I}^{-1}[\mathrm{vec}(\boldsymbol{\Phi})]\boldsymbol{\Xi}_j \cong \mathbf{I}_{m^2}$, $j = 1, \ldots, p$. More precisely, after some algebra it can be checked that

$$\mathbf{P}_{jj} = \ \boldsymbol{\Xi}_j'\mathbf{I}^{-1}[\mathrm{vec}(\boldsymbol{\Phi})]\boldsymbol{\Xi}_j = \mathbf{A}_j \otimes \mathbf{I}_m \ , \tag{5.27}$$

where $\mathbf{A}_j \cong \mathbf{I}_m$, $j = 1, \ldots, p$. The justification of (5.27), that confirms the findings in the *VAR(1)*, *VAR(2)*, and *VAR(3)* models above, is given in appendix 5.1.

An additional $m = 2$ *VAR(2)* model that will be used later is given by the roots in table 5.5. This produces the $2 \times 2$ matrices:

$$\boldsymbol{\Phi}_1 = \begin{pmatrix} 0.2447 & 0.0212 \\ -0.0133 & 0.3031 \end{pmatrix} \quad , \quad \boldsymbol{\Phi}_2 = \begin{pmatrix} -0.0323 & 0.0041 \\ -0.0026 & -0.0210 \end{pmatrix} \ . \tag{5.28}$$

The covariance matrix $\boldsymbol{\Sigma}$ is taken as in (5.4).

| $j$ | $\varsigma_{j,1}$ | $\lvert\varsigma_{j,1}\rvert$ | $\varsigma_{j,2}$ | $\lvert\varsigma_{j,2}\rvert$ |
|---|---|---|---|---|
| 1 | $4.0000 + 4.0000\,i$ | 5.6568 | $4.0000 - 4.0000\,i$ | 5.6568 |
| 2 | $6.1569 + 0.0000\,i$ | 6.1569 | $7.3882 + 0.0000\,i$ | 7.3882 |

Table 5.5: Roots of the determinantal equation $|\boldsymbol{\Phi}(z)| = 0$ of the bivariate *VAR(2)* model (5.28)−(5.4) of section 5.2.2

### 5.2.3 *VMA(q)* models

The generation of *VMA(1)* models can be treated as the *VAR(1)* case. For instance, taking eigenvalues $\delta_j = 0.0901 \pm 0.0433i$, $j = 1, 2$, with $|\delta_1| = |\delta_2| = 0.0999 < 1$; and eigenvectors $\boldsymbol{\gamma}_1 = (2.7071, 0.2768 + 0.3847i)'$, $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (2.7071, 0.2768 - 0.3847i)'$, the identity $\boldsymbol{\Theta}_1 = \mathbf{C}\mathbf{D}\mathbf{C}^{-1}$, where $\mathbf{C} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ and $\mathbf{D} = \mathrm{diag}(\delta_1, \delta_2)$, leads to:

$$\boldsymbol{\Theta}_1 = \begin{pmatrix} 0.0589 & 0.3047 \\ -0.0093 & 0.1212 \end{pmatrix} . \tag{5.29}$$

The covariance matrix of the errors is selected as

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{pmatrix} . \tag{5.30}$$

Expressions (5.29) and (5.30) can be used to form a *VMA(1)* process of the form $\mathbf{X}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Theta}_1\boldsymbol{\varepsilon}_{t-1}$. The roots $\varsigma_{j,1}$, $j = 1, 2$, of the determinantal equation $|\boldsymbol{\Theta}(z)| = |\mathbf{I}_m + \boldsymbol{\Theta}_1 z| = 0$ are related to the eigenvalues in the form $\varsigma_{j,1} = -\delta_j$, $j = 1, 2$.

For the bivariate *VMA(1)* model (5.29)−(5.30), the asymptotic variances of the adjusted residual traces $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 4$ appear in table 5.6 below:

| lag | var |
|---|---|
| 1 | 0.0498 |
| 2 | 0.9516 |
| 3 | 0.9986 |
| 4 | 1.0000 |

Table 5.6: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 5$, in the bivariate *VMA(1)* model (5.29)−(5.30) of section 5.2.3

In this case, only the coefficient $\mathbf{a}_2'\mathbf{P}_{11}\mathbf{a}_2$ is close enough to 1. It can be also checked numerically that

$$\mathbf{P}_{11} = \boldsymbol{\Xi}_1'\mathbf{I}^{-1}[\mathrm{vec}(\boldsymbol{\Theta}_1)]\boldsymbol{\Xi}_1 = \mathbf{I}_2 \otimes \begin{pmatrix} 0.9925 & -0.0243 \\ -0.0243 & 0.9080 \end{pmatrix} =$$

$$= \begin{pmatrix} 0.9925 & -0.0243 & 0.0000 & 0.0000 \\ -0.0243 & 0.9080 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.9925 & -0.0243 \\ 0.0000 & 0.0000 & -0.0243 & 0.9080 \end{pmatrix} \cong \mathbf{I}_4 .$$

The form of this new $\mathbf{P}_{11}$ is a consequence of identity (5.59) in appendix 5.1.

| $j$ | $\varsigma_{j,1}$ | $\lvert\varsigma_{j,1}\rvert$ | $\varsigma_{j,2}$ | $\lvert\varsigma_{j,2}\rvert$ |
|---|---|---|---|---|
| 1 | $2.0000 + 2.0000\,i$ | 2.8284 | $2.0000 - 2.0000\,i$ | 2.8284 |
| 2 | $3.3284 + 0.0000\,i$ | 3.3284 | $3.9941 + 0.0000\,i$ | 3.9941 |

Table 5.7: Roots of the determinantal equation $|\boldsymbol{\Theta}(z)| = 0$ of the bivariate *VMA(2)* model (5.31)−(5.30) of section 5.2.3

The construction of *VMA(q)* processes for $q > 1$ proceeds by adapting conveniently the method of section 5.2.2 for *VAR(p)* models when $p > 1$. For instance, suppose that it is desired to form a bivariate *VMA(2)* model with roots as given in table 5.7. The coefficients $\{d_{j,i}\}$ of the two degree polynomials $p_j(z) = 1 + d_{j,1}z + d_{j,2}z^2$ of (5.16) are determined now from the identities:

$$
\begin{aligned}
d_{j,1} &= -\frac{1}{\varsigma_{j,1}} - \frac{1}{\varsigma_{j,2}}\ ; \\
d_{j,2} &= \frac{1}{\varsigma_{j,1}\varsigma_{j,2}}\ ,
\end{aligned}
$$

$j = 1,\ 2$. The invertible matrix $\mathbf{A}$ of step **(d)** is

$$
\mathbf{A} = \begin{pmatrix} 1.2 & 0.4 \\ 0.2 & 1.0 \end{pmatrix} .
$$

This leads to the $2 \times 2$ matrices:

$$
\boldsymbol{\Theta}_1 = \begin{pmatrix} -0.4964 & -0.0218 \\ 0.0091 & -0.5544 \end{pmatrix} \quad , \quad \boldsymbol{\Theta}_2 = \begin{pmatrix} 0.1286 & -0.0213 \\ 0.0089 & 0.0717 \end{pmatrix} . \tag{5.31}
$$

The covariance matrix of the errors $\boldsymbol{\Sigma}$ is as in expression (5.30). For this *VMA(2)* model $\mathbf{X}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Theta}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Theta}_2\boldsymbol{\varepsilon}_{t-2}$ defined by (5.31)−(5.30), it is found that

$$
\begin{aligned}
\mathbf{P}_{11} &= \mathbf{I}_2 \otimes \begin{pmatrix} 0.9841 & 0.0018 \\ 0.0018 & 0.9841 \end{pmatrix} , \\
\mathbf{P}_{22} &= \mathbf{I}_2 \otimes \begin{pmatrix} 0.7677 & -0.0073 \\ -0.0073 & 0.7116 \end{pmatrix} .
\end{aligned}
$$

The asymptotic variances of the adjusted residual traces $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots,$ 7, are displayed in table 5.8 below:

### 5.2.4 *VARMA(p,q)* models

The construction of *VARMA(p,q)* processes can be done by combining the rules presented earlier in sections 5.2.1, 5.2.2, and 5.2.3 for the autoregressive and moving average parts of the model, respectively.

| lag | var |
|---|---|
| 1 | 0.0108 |
| 2 | 0.2369 |
| 3 | 0.7840 |
| 4 | 0.9882 |
| 5 | 0.9996 |
| 6 | 1.0000 |
| 7 | 1.0000 |

Table 5.8: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 7$, in the bivariate *VMA(2)* model (5.31)−(5.30) of section 5.2.3

For example, a bivariate *VARMA(1,1)* process of the form

$$\mathbf{X}_t - \mathbf{\Phi}_1 \mathbf{X}_{t-1} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} \tag{5.32}$$

can be obtained using the matrices $\mathbf{\Phi}_1$ and $\mathbf{\Theta}_1$ of expressions (5.3) and (5.29) respectively, and the covariance matrix $\mathbf{\Sigma}$ of (5.30). In this case, it is found that

$$\mathbf{P}_{11} = \begin{pmatrix} 0.9986 & -0.0045 & -0.0005 & -0.0016 \\ -0.0045 & 0.9829 & -0.0016 & -0.0062 \\ -0.0005 & -0.0016 & 0.9995 & -0.0017 \\ -0.0016 & -0.0062 & -0.0017 & 0.9934 \end{pmatrix},$$

and

$$\mathbf{P}_{22} = \begin{pmatrix} 0.8680 & 0.0549 & -0.0454 & 0.0654 \\ 0.0549 & 0.7906 & -0.0070 & -0.0397 \\ -0.0454 & -0.0070 & 0.9683 & 0.0451 \\ 0.0654 & -0.0397 & 0.0451 & 0.8948 \end{pmatrix}.$$

Both matrices above are reasonably close to the identity $\mathbf{I}_4$. However, they lack the patterns (5.27) and (5.59) observed in the *VAR(2)* and *VMA(2)* cases, respectively. The asymptotic variances of the adjusted residual traces $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots,$ 7, are displayed in table 5.9 below. According to this numerical information, only the first two scaled adjusted residual traces have a variance markedly smaller than 1. This phenomenon can be explained by adapting adequately the arguments of appendix 5.1. For conciseness, details are omitted.

By considering for the determinantal equation $|\mathbf{\Phi}(z)| = |\mathbf{I}_m - \mathbf{\Phi}_1 z - \mathbf{\Phi}_2 z^2| = 0$ the roots $\varsigma_{1,1} = 5.9999 + 5.9999i$; $\varsigma_{1,2} = 5.9999 - 5.9999i$, so that $|\varsigma_{1,1}| = |\varsigma_{1,2}| = 8.4851$; $\varsigma_{2,1} = 8.9853 + 0.0000i$; and $\varsigma_{2,2} = 10.7823 + 0.0000i$; and using the invertible matrix

$$\mathbf{A} = \begin{pmatrix} 1.2 & 0.4 \\ 0.2 & 1.0 \end{pmatrix},$$

it is obtained

| lag | var |
|---|---|
| 1 | 0.0056 |
| 2 | 0.0532 |
| 3 | 0.9520 |
| 4 | 0.9906 |
| 5 | 0.9988 |
| 6 | 0.9998 |
| 7 | 1.0000 |

Table 5.9: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 7$, in the bivariate *VARMA(1,1)* model (5.32)−(5.30) of section 5.2.4

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.1640 & 0.0160 \\ -0.0067 & 0.2067 \end{pmatrix} \quad , \quad \mathbf{\Phi}_2 = \begin{pmatrix} -0.0141 & 0.0015 \\ -0.0006 & -0.0101 \end{pmatrix} \; . \tag{5.33}$$

Taking in turn roots $\varsigma_{1,1} = 4.0000 + 4.0000i$; $\varsigma_{1,2} = 4.0000 - 4.0000i$, with $|\varsigma_{1,1}| = |\varsigma_{1,2}| = 5.6568$; $\varsigma_{2,1} = 6.1569 + 0.0000i$; and $\varsigma_{2,2} = 7.3882 + 0.0000i$ for the equation $|\mathbf{\Theta}(z)| = |\mathbf{I}_m + \mathbf{\Theta}_1 z + \mathbf{\Theta}_2 z^2| = 0$, and using the same $\mathbf{A}$ as above it is found that

$$\mathbf{\Theta}_1 = \begin{pmatrix} -0.2466 & -0.0205 \\ 0.0085 & -0.3012 \end{pmatrix} \quad , \quad \mathbf{\Theta}_2 = \begin{pmatrix} 0.0319 & -0.0040 \\ 0.0017 & 0.0213 \end{pmatrix} \; . \tag{5.34}$$

The $2 \times 2$ matrices of (5.33) and (5.34), combined with the error covariance matrix $\mathbf{\Sigma}$ of (5.30), lead finally to a *VARMA(2,2)* process of the form

$$\mathbf{X}_t - \mathbf{\Phi}_1 \mathbf{X}_{t-1} - \mathbf{\Phi}_2 \mathbf{X}_{t-2} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} + \mathbf{\Theta}_2 \boldsymbol{\varepsilon}_{t-1} \; . \tag{5.35}$$

## 5.3 Behavior of the adjusted traces

This section studies and compares the properties of the different versions of the adjusted traces: the residual $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ of (1.29); the modified $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ of (4.11); and the model error version $\mathrm{tr}(\mathbf{R}_k)/\sqrt{m}$ of (2.43).

As obtained in expression (2.39) of section 2.5,

$$\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m} \overset{D}{\cong} N(0, 1 - \mathbf{a}'_m \mathbf{P}_{kk} \mathbf{a}_m) \; ,$$

where $\mathbf{a}_m = \mathrm{vec}(\mathbf{I}_m)/\sqrt{m}$ is a unit $m^2 \times 1$ vector, and $\mathbf{P}_{kk} = \mathbf{\Xi}'_k \mathbf{I}^{-1}(\mathbf{\Lambda}) \mathbf{\Xi}_k$ is the $m^2 \times m^2$ matrix defined in (3.6), $k = 1, \ldots, M$. The bands of (2.41),

$$\pm 1.96\, n^{-1/2} \, (1 - \mathbf{a}'_m \mathbf{P}_{kk} \mathbf{a}_m)^{1/2} \; , \quad 1 \le k \le M \; ,$$

corresponding to a sample size of $n = 250$ and a nominal level $\alpha = .05$, are plotted in figure 5.3 for seven bivariate time series models:

**(a)** The *VAR(1)* of expression (5.2) of section 5.2.1;

**(b)** The *VAR(2)* of expressions (5.28)–(5.4) in section 5.2.2;

**(c)** The *VAR(3)* of expressions (5.24)–(5.4) in section 5.2.2;

**(d)** The *VMA(1)* of expressions (5.29)–(5.30) of section 5.2.3;

**(e)** The *VMA(2)* of expressions (5.31)–(5.30) of section 5.2.3;

**(f)** The *VARMA(1,1)* of expression (5.32) in section 5.2.4; and

**(g)** The *VARMA(2,2)* of expression (5.35) in section 5.2.4.

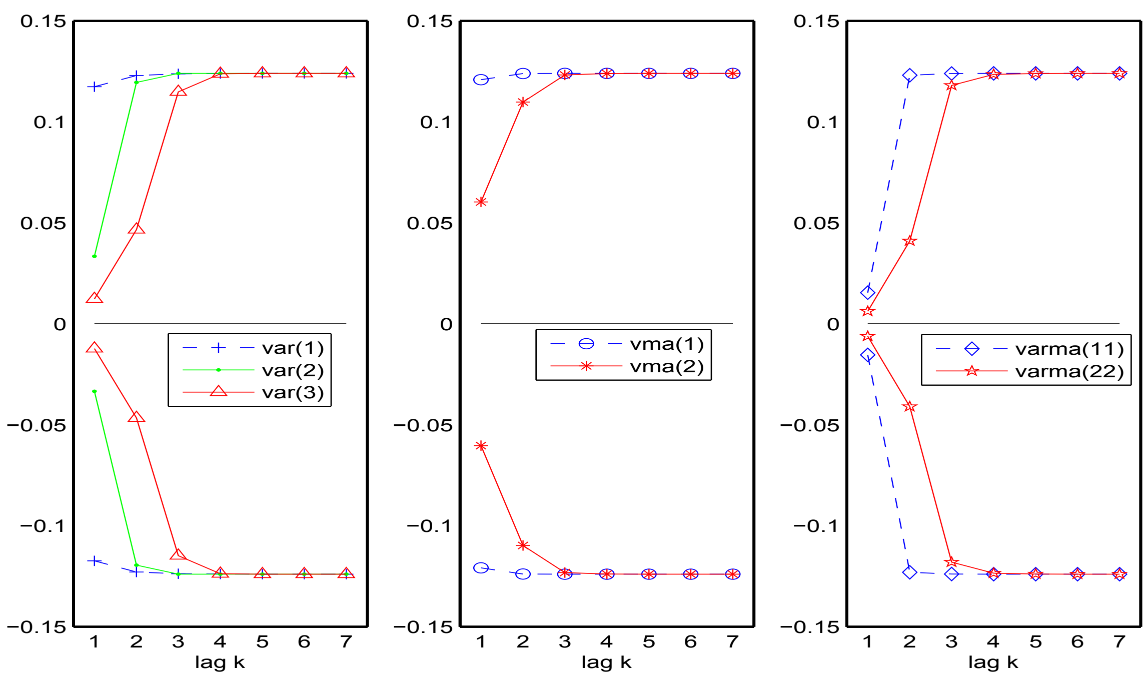


Figure 5.3: Bands $\pm 1.96 n^{-1/2}(1 - \mathbf{a}_m' \mathbf{P}_{kk} \mathbf{a}_m)^{1/2}$, $k = 1, \ldots, M$, with $n = 250$ for the seven models of section 5.3

The message of figure 5.3 is confirmed numerically by the values of the asymptotic variances $1 - \mathbf{a}_m' \mathbf{P}_{kk} \mathbf{a}_m$ of (2.40), $k = 1, \ldots, M$. These appear in table 5.10 below:

The sizes of the entries of table 5.10 agree with the numerical findings of appendix 5.1. A real data application of the diagnostic check provided by the bands of figure 5.3 will be given later in section 5.7.

| lag | *VAR(1)* | *VAR(2)* | *VAR(3)* | *VMA(1)* | *VMA(2)* | *VARMA(1,1)* | *VARMA(2,2)* |
|---|---|---|---|---|---|---|---|
| 1 | 0.1228 | 0.0007 | 0.0016 | 0.0498 | 0.0108 | 0.0056 | 0.0001 |
| 2 | 0.8964 | 0.0727 | 0.0099 | 0.9516 | 0.2369 | 0.0532 | 0.0001 |
| 3 | 0.9837 | 0.9293 | 0.1413 | 0.9986 | 0.7840 | 0.9520 | 0.0002 |
| 4 | 0.9975 | 0.9999 | 0.8573 | 1.0000 | 0.9882 | 0.9906 | 0.1258 |
| 5 | 0.9997 | 1.0000 | 0.9980 | – | 0.9996 | 0.9988 | 0.9166 |
| 6 | 1.0000 | – | 0.9999 | – | 1.0000 | 0.9998 | 0.9582 |
| 7 | – | – | 1.0000 | – | 1.0000 | 1.0000 | 0.9985 |
| 8 | – | – | – | – | – | – | 0.9999 |

Table 5.10: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, M$, for the seven models of section 5.3

On the other hand, in expression (4.13) of chapter 4 it was found that

$$\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m} \overset{D}{\cong} N(0,1) \ , \quad k \geq p+q+1 \ .$$

In turn, from convergence (2.47) in Proposition 2.6.2 it also follows that

$$\sqrt{n}\,\mathrm{tr}(\mathbf{R}_k)/\sqrt{m} \overset{D}{\cong} N(0,1) \ , \quad k \geq 1 \ .$$

In order to compare the behavior of the different adjusted traces, $N = 1000$ independent replicas of size $n = 250$ are generated from the bivariate *VAR(1)*, *VMA(1)*, and *VARMA(1,1)* models considered in figure 5.3 and table 5.10. The results are presented in the histograms that appear in figures 5.4, 5.5, and 5.6, respectively. For ease of presentation, an axis label `tr1` refers to the rescaled adjusted residual trace $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_1)/\sqrt{m}$ of (1.29); a label `ts1` to the modified $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{S}}_{p+q+1})/\sqrt{m}$ of (4.11); and a label `te1` to the model error version $\sqrt{n}\,\mathrm{tr}(\mathbf{R}_1)/\sqrt{m}$ of (2.43).

The graphical patterns of these histograms support the accuracy of the the theoretical results of chapters 2, 3, and 4 on the behavior of the adjusted traces for a large enough sample size $n$. In general, both $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ in (4.11) and $\sqrt{n}\,\mathrm{tr}(\mathbf{R}_k)/\sqrt{m}$ in (2.43) are close to a $N(0,1)$. On the other hand, for low values of the lag k, the rescaled adjusted residual traces $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ of (1.29) behave like a centered normal with a variance smaller than 1. The variance goes to 1 when $k$ increases, and thus the original adjusted residual traces will tend to behave as the modified ones.

The aforementioned analogy holds not only in distribution, but also numerically. This is because, as seen in (4.32), for $k$ and $n$ large enough

$$\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m} = \mathbf{a}'_m \mathrm{vec}(\widehat{\mathbf{S}}_k) \cong \mathbf{a}'_m \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_k\widehat{\boldsymbol{\Sigma}}^{-1/2}) = \mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m} \ .$$

Table 5.11 displays the values of the statistics $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ and $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ for the last generated sample of the simulation experiment that produces the histograms in

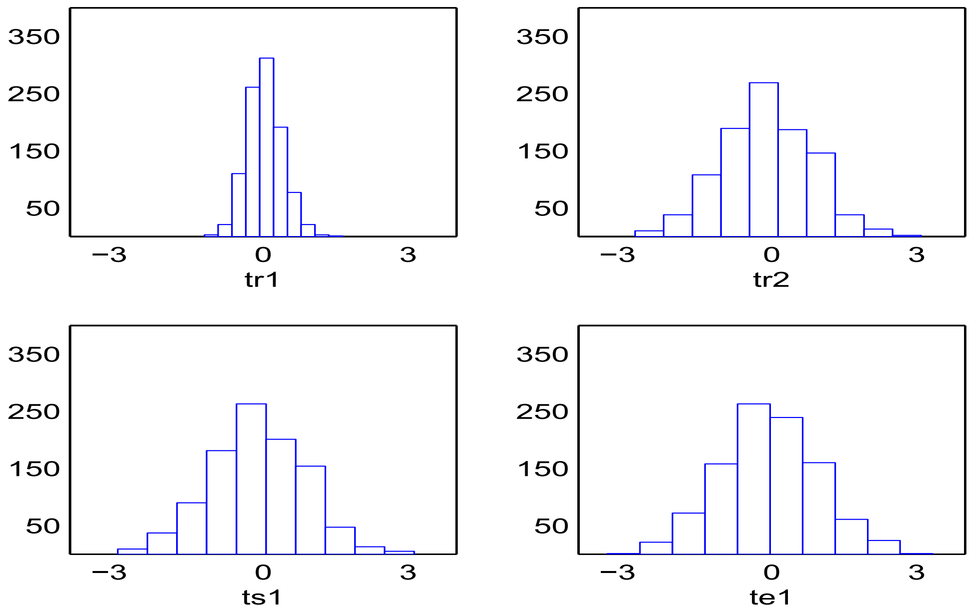


Figure 5.4: Histograms of the adjusted traces for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VAR(1)* (5.2) of section 5.2.1

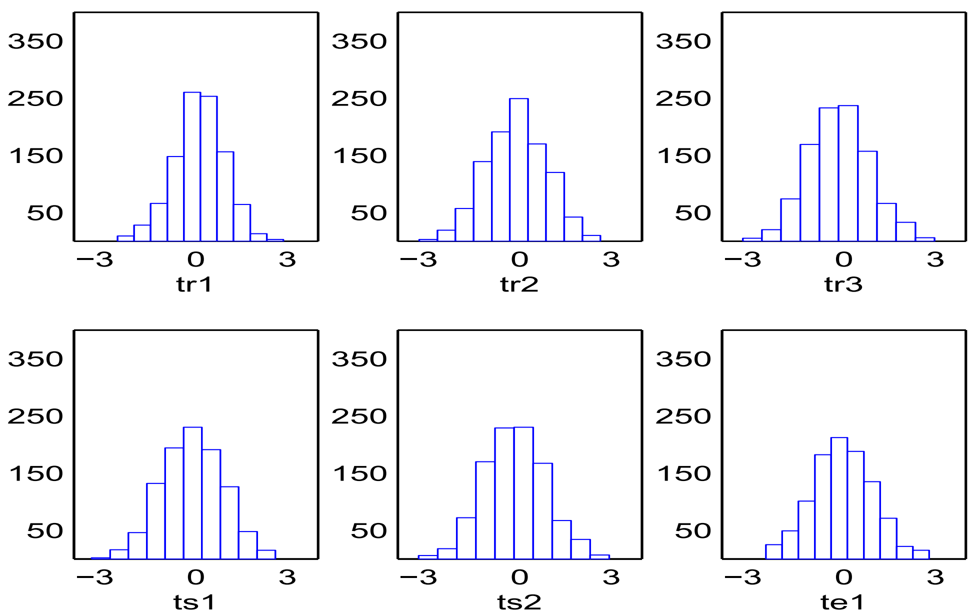


Figure 5.5: Histograms of the adjusted traces for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VMA(1)* (5.29)−(5.30) of section 5.2.3

figures 5.4, 5.5, and 5.6. Only the values of the adjusted traces that are different are presented. From table 5.11, just the first traces differ to some extent. As it may be expected, the differences, that are always moderate in size, increase with the

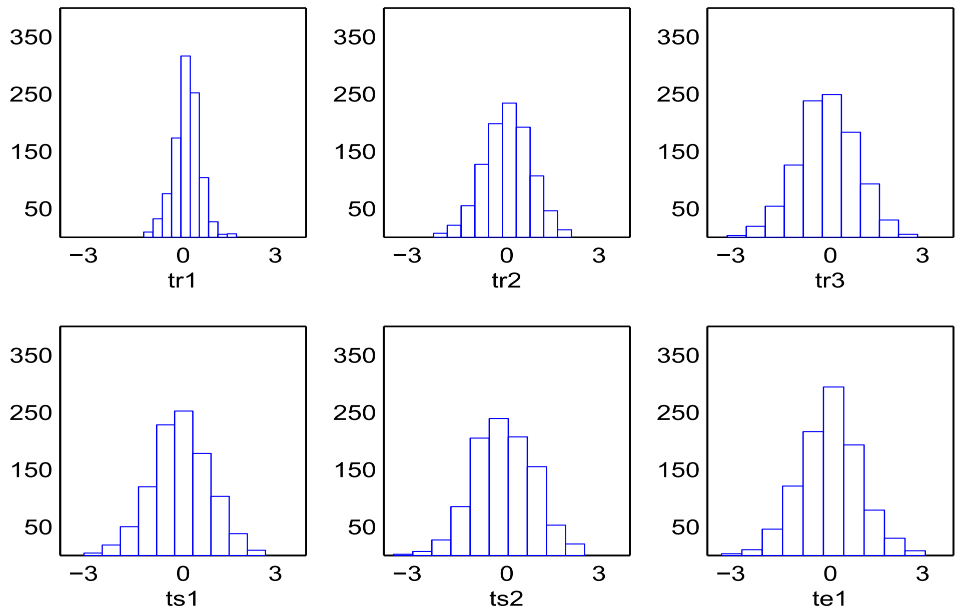


Figure 5.6: Histograms of the adjusted traces for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VARMA(1,1)* (5.32) in section 5.2.4

complexity of the model.

| | *VAR(1)* | | *VMA(1)* | | *VARMA(1,1)* | |
|---|---|---|---|---|---|---|
| lag | $\frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}$ | $\frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}$ | $\frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}$ | $\frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}$ | $\frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}$ | $\frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}$ |
| 1 | 0.0037 | – | −0.0370 | – | −0.0002 | – |
| 2 | 0.0908 | 0.0754 | −0.0345 | −0.0479 | −0.0033 | – |
| 3 | −0.0569 | −0.0630 | 0.1574 | 0.1565 | −0.1090 | −0.1047 |
| 4 | −0.0524 | −0.0531 | −0.1706 | −0.1700 | 0.0546 | 0.0505 |
| 5 | −0.0896 | −0.0896 | −0.0837 | −0.0837 | 0.1633 | 0.1652 |
| 6 | ... | ... | ... | ... | −0.1800 | −0.1803 |
| 7 | ... | ... | ... | ... | −0.0009 | −0.0007 |
| 8 | ... | ... | ... | ... | −0.0320 | −0.0320 |

Table 5.11: Original and modified adjusted residual traces for the last generated sample of size $n = 250$ in the experiment underlying the histograms appearing in figures 5.4, 5.5, and 5.6

## 5.4 Comparisons between goodness-of-fit processes

This section compares the behavior in size and power of the functionals $KS$ of (4.47); $CVM$ of (4.46); and $PCVM$ of (4.48), when they are used with the different goodness-of-fit processes studied in this thesis: the error process $\{W_n^m(u) : 0 \leq u \leq 1\}$ of (2.44); the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (1.28); and the modified process

$\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$ of (1.30). The rejection criteria are of the form (3.34)−(4.44). Specifically, for the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (1.28) criterion $KS$ rejects for a nominal level $\alpha$ the adequacy of a given model when

$$KS = \sup_{1 \leq j \leq n} |\widehat{W}_n^m(j/n)| > KS_\alpha ,$$

where $KS_\alpha$ is the proper quantile of the distribution of $\sup_{1 \leq u \leq 1} |B(u)|$. Similarly for criteria $CVM$ and $PCVM$. For the error process $\{W_n^m(u) : 0 \leq u \leq 1\}$ of (2.44), application of these functionals above can be seen as a verification study of the convergence result of Theorem 2.6.2.

### 5.4.1 Size

For the bivariate *VAR(1)*, *VMA(1)*, and *VARMA(1,1)* models considered in figure 5.3; table 5.10; and histograms 5.4−5.5−5.6, $N = 1000$ independent replicas of size $n = 250$ are generated. Nominal significance levels considered are $\alpha = .1$, .05, and .01. Results are presented in table 5.12 for *VAR(1)*; table 5.13 for *VMA(1)*; and table 5.14 for *VARMA(1,1)*. For a given value of $\alpha$, the information contained in the columns of these tables is as follows:

**(a)** Empirical proportion of rejections, say $\widehat{p}_N$;

**(b)** Lower bound of a 95% confidence interval for the true probability of rejection at level $\alpha$, $\widehat{p}_N - 1.96\sqrt{\widehat{p}_N(1-\widehat{p}_N)/N}$;

**(c)** Upper bound version of **(b)**, $\widehat{p}_N + 1.96\sqrt{\widehat{p}_N(1-\widehat{p}_N)/N}$;

**(d)** Theoretical quantile of the specific criterion;

**(e)** Empirical quantile for the $N = 1000$ replicas simulated.

As a conclusion from tables 5.12−5.13−5.14, the empirical size of the criteria based on functionals applied on the error and modified processes is quite close to the nominal. This justifies the convergence results in Theorems 2.6.2 and 4.4.1, respectively. As expected, the behavior of $CVM$ and $PCVM$ is almost identical. Consequently, $PCVM$ will be ignored from now on. However, when considering $KS$, $CVM$, and $PCVM$ on the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (1.28), the size of the regions of the form (3.34),

$$H[\widehat{W}_n^m(u)] \geq H_\alpha[B(u)] ,$$

is well below $\alpha$. This indicates that inequality (5.52), consequence in turn of the result (3.32) by Anderson (1955), can be very severe in practice.

### 5.4.2 Power

For making power comparisons between the functionals applied on the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (1.28) and the modified process $\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$ of (1.30), a bivariate *VARMA(1,1)* model of the form $\mathbf{X}_t - \mathbf{\Phi}_1 \mathbf{X}_{t-1} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1(\beta)\boldsymbol{\varepsilon}_{t-1}$ is considered, where $\mathbf{\Theta}_1(\beta) = \beta\mathbf{\Theta}_1$; $\beta$ is a real parameter in the interval $[0, 1)$;

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.1868 & 0.1787 \\ -0.0122 & 0.2101 \end{pmatrix} ; \tag{5.36}$$

and

$$\mathbf{\Theta}_1 = \begin{pmatrix} 1.3008 & -0.1945 \\ 1.7858 & 0.5017 \end{pmatrix} . \tag{5.37}$$

The matrix $\mathbf{\Phi}_1$ of (5.36) is obtained by taking eigenvalues $\delta_j = 0.1984 \pm 0.0452\,i$, $j = 1,\ 2$, so that $|\delta_1| = |\delta_2| = 0.2035 < 1$. The associated eigenvectors are $\boldsymbol{\gamma}_1 = (0.9675, 0.0632 + 0.2447\,i)'$; $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (0.9675, 0.0632 - 0.2447\,i)'$. The eigenvalues of the matrix $\mathbf{\Theta}_1$ of (5.37) are $\delta_j = 0.9013 \pm 0.4333\,i$, $j = 1,\ 2$, so that $|\delta_1| = |\delta_2| = 1.0000$. The eigenvectors are $\boldsymbol{\gamma}_1 = (0.2125 + 0.2304\,i, 0.9496)'$, and $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (0.2125 - 0.2304\,i, 0.9496)'$. The covariance matrix $\mathbf{\Sigma}$ of the errors is as in (5.30).

For each value in a grid of values of the parameter $0 \leq \beta < 1$, $N = 1000$ independent data samples of length $n = 200$ are generated from the process $\mathbf{X}_t - \mathbf{\Phi}_1 \mathbf{X}_{t-1} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1(\beta)\boldsymbol{\varepsilon}_{t-1}$ defined by the matrices in (5.36)–(5.37)–(5.30). If a *VAR(1)* process is postulated and fitted, the value of $\beta = 0$ corresponds to the null *VAR(1)* model. The values of $0 < \beta < 1$ define an alternative *VARMA(1,1)* model. Thus, the plots of the corresponding empirical proportions of rejections at level $\alpha = .05$ versus $\beta$ give a graphical display of the power function of the method. The modulus of the eigenvalues of $\mathbf{\Theta}_1(\beta)$ is $\beta$, so that when $\beta \to 1$ the alternative *VARMA(1,1)* process is close to having a unit root. The results are displayed at the left part of figure 5.7. For moderate values of $\beta$, both $KS$ and $CVM$ on the modified process are more powerful than the same functionals on the original process. All the power functions tend to unity when $\beta \to 1$. The larger power of $KS$ and $CVM$ for moderate to large values of $\beta$ in the original residual process can be explained by the fact that the $m \times m$ residual correlation matrices $\widehat{\mathbf{R}}_k$ are based on residual vectors $\widehat{\boldsymbol{\varepsilon}}_t$ that are more sensitive to departures form the null assumption than the modified matrices $\widehat{\mathbf{S}}_k$.

An additional experiment is conducted. A $m = 2$ *VARMA(2,2)* process of the form $\mathbf{X}_t - \mathbf{\Phi}_1 \mathbf{X}_{t-1} - \mathbf{\Phi}_2 \mathbf{X}_{t-2} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1(\beta)\boldsymbol{\varepsilon}_{t-1} + \mathbf{\Theta}_2(\beta)\boldsymbol{\varepsilon}_{t-2}$ is considered, where $0 \leq \beta < 1$. The $2 \times 2$ matrices $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are obtained with the method of section

| $\alpha = .10$ | **(a)** | **(b)** | **(c)** | **(d)** | **(e)** |
|---|---|---|---|---|---|
| | Error process | | | | |
| KS | 0.0650 | 0.0497 | 0.0803 | 1.2238 | 1.1261 |
| CVM | 0.0990 | 0.0805 | 0.1175 | 0.3473 | 0.3395 |
| PCVM | 0.0990 | 0.0805 | 0.1175 | 0.3473 | 0.3395 |
| | Residual process | | | | |
| KS | 0.0030 | -0.0004 | 0.0064 | 1.2238 | 0.8191 |
| CVM | 0.0000 | 0.0000 | 0.0000 | 0.3473 | 0.1369 |
| PCVM | 0.0000 | 0.0000 | 0.0000 | 0.3473 | 0.1369 |
| | Modified process | | | | |
| KS | 0.0860 | 0.0686 | 0.1034 | 1.2238 | 1.1864 |
| CVM | 0.1090 | 0.0897 | 0.1283 | 0.3473 | 0.3494 |
| PCVM | 0.1090 | 0.0897 | 0.1283 | 0.3473 | 0.3494 |
| $\alpha = .05$ | **(a)** | **(b)** | **(c)** | **(d)** | **(e)** |
| | Error process | | | | |
| KS | 0.0330 | 0.0219 | 0.0441 | 1.3582 | 1.2549 |
| CVM | 0.0490 | 0.0356 | 0.0624 | 0.4614 | 0.4476 |
| PCVM | 0.0490 | 0.0356 | 0.0624 | 0.4614 | 0.4476 |
| | Residual process | | | | |
| KS | 0.0000 | 0.0000 | 0.0000 | 1.3582 | 0.9343 |
| CVM | 0.0000 | 0.0000 | 0.0000 | 0.4614 | 0.1641 |
| PCVM | 0.0000 | 0.0000 | 0.0000 | 0.4614 | 0.1641 |
| | Modified process | | | | |
| KS | 0.0390 | 0.0270 | 0.0510 | 1.3582 | 1.3284 |
| CVM | 0.0440 | 0.0313 | 0.0567 | 0.4614 | 0.4443 |
| PCVM | 0.0440 | 0.0313 | 0.0567 | 0.4614 | 0.4443 |
| $\alpha = .01$ | **(a)** | **(b)** | **(c)** | **(d)** | **(e)** |
| | Error process | | | | |
| KS | 0.0040 | 0.0001 | 0.0079 | 1.6277 | 1.5165 |
| CVM | 0.0050 | 0.0006 | 0.0094 | 0.7435 | 0.6937 |
| PCVM | 0.0050 | 0.0006 | 0.0094 | 0.7435 | 0.6937 |
| | Residual process | | | | |
| KS | 0.0000 | 0.0000 | 0.0000 | 1.6277 | 1.0197 |
| CVM | 0.0000 | 0.0000 | 0.0000 | 0.7435 | 0.2392 |
| PCVM | 0.0000 | 0.0000 | 0.0000 | 0.7435 | 0.2392 |
| | Modified process | | | | |
| KS | 0.0150 | 0.0075 | 0.0225 | 1.6277 | 1.6702 |
| CVM | 0.0090 | 0.0031 | 0.0149 | 0.7435 | 0.7081 |
| PCVM | 0.0090 | 0.0031 | 0.0149 | 0.7435 | 0.7081 |

Table 5.12: Empirical sizes for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VAR(1)* (5.2) of section 5.2.1

| $\alpha = .10$ | **(a)** | **(b)** | **(c)** | **(d)** | **(e)** |
|---|---|---|---|---|---|
| | Error process | | | | |
| KS | 0.0660 | 0.0506 | 0.0814 | 1.2238 | 1.1440 |
| CVM | 0.1090 | 0.0897 | 0.1283 | 0.3473 | 0.3619 |
| PCVM | 0.1090 | 0.0897 | 0.1283 | 0.3473 | 0.3619 |
| | Residual process | | | | |
| KS | 0.0350 | 0.0236 | 0.0464 | 1.2238 | 1.0150 |
| CVM | 0.0420 | 0.0296 | 0.0544 | 0.3473 | 0.2364 |
| PCVM | 0.0420 | 0.0296 | 0.0544 | 0.3473 | 0.2364 |
| | Modified process | | | | |
| KS | 0.0840 | 0.0668 | 0.1012 | 1.2238 | 1.1801 |
| CVM | 0.0870 | 0.0695 | 0.1045 | 0.3473 | 0.3321 |
| PCVM | 0.0870 | 0.0695 | 0.1045 | 0.3473 | 0.3321 |
| $\alpha = .05$ | **(a)** | **(b)** | **(c)** | **(d)** | **(e)** |
| | Error process | | | | |
| KS | 0.0270 | 0.0170 | 0.0370 | 1.3582 | 1.2738 |
| CVM | 0.0520 | 0.0382 | 0.0658 | 0.4614 | 0.4621 |
| PCVM | 0.0520 | 0.0382 | 0.0658 | 0.4614 | 0.4621 |
| | Residual process | | | | |
| KS | 0.0120 | 0.0053 | 0.0187 | 1.3582 | 1.1503 |
| CVM | 0.0170 | 0.0090 | 0.0250 | 0.4614 | 0.3162 |
| PCVM | 0.0170 | 0.0090 | 0.0250 | 0.4614 | 0.3162 |
| | Modified process | | | | |
| KS | 0.0430 | 0.0304 | 0.0556 | 1.3582 | 1.3398 |
| CVM | 0.0450 | 0.0322 | 0.0578 | 0.4614 | 0.4455 |
| PCVM | 0.0450 | 0.0322 | 0.0578 | 0.4614 | 0.4455 |
| $\alpha = .01$ | **(a)** | **(b)** | **(c)** | **(d)** | **(e)** |
| | Error process | | | | |
| KS | 0.0070 | 0.0018 | 0.0122 | 1.6277 | 1.5704 |
| CVM | 0.0120 | 0.0053 | 0.0187 | 0.7435 | 0.7800 |
| PCVM | 0.0120 | 0.0053 | 0.0187 | 0.7435 | 0.7800 |
| | Residual process | | | | |
| KS | 0.0020 | -0.0008 | 0.0048 | 1.6277 | 1.3752 |
| CVM | 0.0020 | -0.0008 | 0.0048 | 0.7435 | 0.5249 |
| PCVM | 0.0020 | -0.0008 | 0.0048 | 0.7435 | 0.5249 |
| | Modified process | | | | |
| KS | 0.0110 | 0.0045 | 0.0175 | 1.6277 | 1.6528 |
| CVM | 0.0090 | 0.0031 | 0.0149 | 0.7435 | 0.6978 |
| PCVM | 0.0090 | 0.0031 | 0.0149 | 0.7435 | 0.6978 |

Table 5.13: Empirical sizes for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VMA(1)* (5.29)−(5.30) of section 5.2.3

| $\alpha = .10$ | **(a)** | **(b)** | **(c)** | **(d)** | **(e)** |
|---|---|---|---|---|---|
| | Error process | | | | |
| KS | 0.0570 | 0.0426 | 0.0714 | 1.2238 | 1.1314 |
| CVM | 0.0930 | 0.0750 | 0.1110 | 0.3473 | 0.3331 |
| PCVM | 0.0930 | 0.0750 | 0.1110 | 0.3473 | 0.3331 |
| | Residual process | | | | |
| KS | 0.0030 | -0.0004 | 0.0064 | 1.2238 | 0.8145 |
| CVM | 0.0020 | -0.0008 | 0.0048 | 0.3473 | 0.1235 |
| PCVM | 0.0020 | -0.0008 | 0.0048 | 0.3473 | 0.1235 |
| | Modified process | | | | |
| KS | 0.0770 | 0.0605 | 0.0935 | 1.2238 | 1.1613 |
| CVM | 0.0780 | 0.0614 | 0.0946 | 0.3473 | 0.3122 |
| PCVM | 0.0780 | 0.0614 | 0.0946 | 0.3473 | 0.3122 |
| $\alpha = .05$ | **(a)** | **(b)** | **(c)** | **(d)** | **(e)** |
| | Error process | | | | |
| KS | 0.0300 | 0.0194 | 0.0406 | 1.3582 | 1.2356 |
| CVM | 0.0490 | 0.0356 | 0.0624 | 0.4614 | 0.4419 |
| PCVM | 0.0490 | 0.0356 | 0.0624 | 0.4614 | 0.4419 |
| | Residual process | | | | |
| KS | 0.0000 | 0.0000 | 0.0000 | 1.3582 | 0.8871 |
| CVM | 0.0000 | 0.0000 | 0.0000 | 0.4614 | 0.1470 |
| PCVM | 0.0000 | 0.0000 | 0.0000 | 0.4614 | 0.1470 |
| | Modified process | | | | |
| KS | 0.0330 | 0.0219 | 0.0441 | 1.3582 | 1.2903 |
| CVM | 0.0420 | 0.0296 | 0.0544 | 0.4614 | 0.4203 |
| PCVM | 0.0420 | 0.0296 | 0.0544 | 0.4614 | 0.4203 |
| $\alpha = .01$ | **(a)** | **(b)** | **(c)** | **(d)** | **(e)** |
| | Error process | | | | |
| KS | 0.0060 | 0.0012 | 0.0108 | 1.6277 | 1.5429 |
| CVM | 0.0100 | 0.0038 | 0.0162 | 0.7435 | 0.7364 |
| PCVM | 0.0100 | 0.0038 | 0.0162 | 0.7435 | 0.7364 |
| | Residual process | | | | |
| KS | 0.0000 | 0.0000 | 0.0000 | 1.6277 | 1.0106 |
| CVM | 0.0000 | 0.0000 | 0.0000 | 0.7435 | 0.2202 |
| PCVM | 0.0000 | 0.0000 | 0.0000 | 0.7435 | 0.2202 |
| | Modified process | | | | |
| KS | 0.0060 | 0.0012 | 0.0108 | 1.6277 | 1.5535 |
| CVM | 0.0070 | 0.0018 | 0.0122 | 0.7435 | 0.6533 |
| PCVM | 0.0070 | 0.0018 | 0.0122 | 0.7435 | 0.6533 |

Table 5.14: Empirical sizes for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VARMA(1,1)* (5.32) in section 5.2.4

5.2.2. The roots of the determinantal equation $|\mathbf{\Phi}(z)| = |\mathbf{I}_2 - \mathbf{\Phi}_1 z - \mathbf{\Phi}_2 z^2| = 0$ are given by $\varsigma_{1,1} = 0.7589 + 1.5178\,i$, $\varsigma_{1,2} = 0.7589 - 1.5178\,i$, so that $|\varsigma_{1,i}| = 1.6979$, $i = 1$, 2; and $\varsigma_{2,1} = 2.1971 + 0.0000\,i$, $\varsigma_{2,2} = 2.8562 + 0.0000\,i$. The invertible matrix

$$\mathbf{A} = \begin{pmatrix} 1.8 & 0.7 \\ 0.2 & 1.0 \end{pmatrix}$$

leads to

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.5036 & 0.2112 \\ -0.0335 & 0.8287 \end{pmatrix} \quad , \quad \mathbf{\Phi}_2 = \begin{pmatrix} -0.3631 & 0.1426 \\ -0.0226 & -0.1435 \end{pmatrix} . \tag{5.38}$$

On the other hand, the matrices $\mathbf{\Theta}_1(\beta)$ and $\mathbf{\Theta}_2(\beta)$ are selected so that $\mathbf{\Theta}_1(0) = \mathbf{\Theta}_2(0) = \mathbf{0}_{2\times 2}$. For $0 < \beta < 1$, the roots of the determinantal equation $|\mathbf{I}_2 + \mathbf{\Theta}_1(\beta) z + \mathbf{\Theta}_2(\beta) z^2| = 0$ are of the form

$$\varsigma_{j,i}(\beta) = q(\beta)\,\varsigma_{j,i} \ , \quad j, i = 1, 2 \ , \tag{5.39}$$

where $q(\beta) = 2.8469 - 1.8469\,\beta$ for $0 < \beta < 1$. The $\varsigma_{j,i}$ are as specified in table 5.15 below. The covariance matrix $\mathbf{\Sigma}$ of the errors is taken again as in (5.30). The invertible matrix of the method of section 5.2.2 is now

$$\mathbf{A} = \begin{pmatrix} 2.2 & 0.4 \\ 0.2 & 1.0 \end{pmatrix} .$$

| $j$ | $\varsigma_{j,1}$ | $\lvert\varsigma_{j,1}\rvert$ | $\varsigma_{j,2}$ | $\lvert\varsigma_{j,2}\rvert$ |
|---|---|---|---|---|
| 1 | $0.7071 + 0.7071\,i$ | 1.0000 | $0.7071 - 0.7071\,i$ | 1.0000 |
| 2 | $1.1095 + 0.0000\,i$ | 1.1095 | $1.1314 + 0.0000\,i$ | 1.1314 |

Table 5.15: Multiples of the roots of the *VMA* part in the parametric bivariate *VARMA(2,2)* model (5.38)−(5.39)−(5.30) of section 5.4.2

Notice that $q(\beta)$ decreases towards 1 when $\beta \to 1$. Thus, the $VMA$ part of the bivariate *VARMA(2,2)* model defined by (5.38)−(5.39)−(5.30) approaches also to a unit root situation. If a *VAR(2)* is postulated and fitted, the value of $\beta = 0$ corresponds to the null *VAR(2)* process. The values of $0 < \beta < 1$ form an alternative *VARMA(2,2)* model. As before, for each value of a grid of values of $0 \le \beta < 1$, $N = 1000$ independent data samples of length $n = 200$ are generated from the *VARMA(2,2)* model (5.38)−(5.39)−(5.30). The associated plot of empirical powers at level $\alpha = .05$ is given at the right of figure 5.7, in which the functionals based on the modified process are clearly much more powerful.

The results of this section confirm the conjectures of section 4.5 relative to both the behavior in size and power of the rejection regions (4.44) and (3.34), $H[\widehat{Z}_n^m(u)] \ge H_\alpha[B(u)]$ and $H[\widehat{W}_n^m(u)] \ge H_\alpha[B(u)]$, respectively.

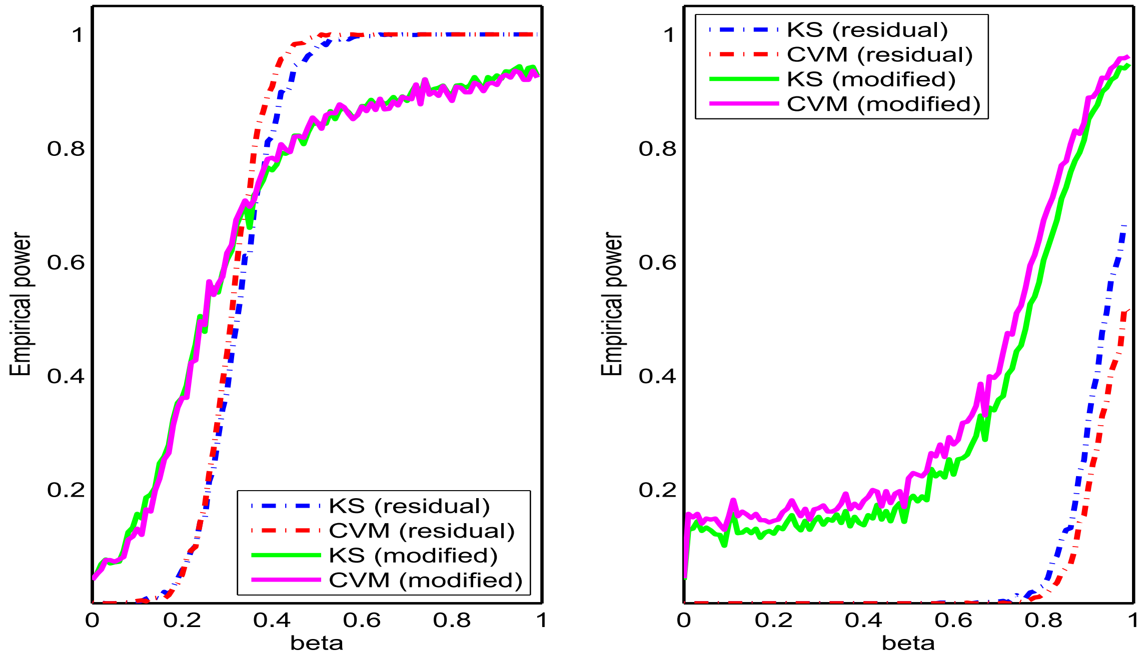


Figure 5.7: Comparison of empirical powers of the residual and modified processes in the two simulation experiments of section 5.4. Left: *VAR(1)* fitted under a parametric *VARMA(1,1)* model; Right: *VAR(2)* fitted under a parametric *VARMA(2,2)* model

## 5.5 Comparisons with previous criteria

As seen in table 5.11, the adjusted residual trace $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ of (1.29), and the corresponding modified version $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ of (4.11) are virtually identical for large enough values of the lag $k$. Thus, it seems adequate to consider a truncated version $\{\widehat{Z}^m_{n,M}(u) : 0 \leq u \leq 1\}$ of the modified process $\{\widehat{Z}^m_n(u) : 0 \leq u \leq 1\}$ of (1.30), where the random function $\widehat{Z}^m_n(u)$ is replaced by

$$\widehat{Z}^m_{n,M}(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\left[\sum_{k=p+q+1}^{M} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}\frac{\sin(K\pi u)}{K} + \sum_{k=M+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(K\pi u)}{K}\right] , \quad (5.40)$$

for an adequately chosen value of $M \geq p+q+1$. This idea was also explored in the univariate case by Ubierna and Velilla (2007, section 4.1).

As such, goodness-of-fit functionals of the form

$$CVT = \int_0^1 [\widehat{Z}^m_{n,M}(u)]^2 du = \frac{n}{m\,\pi^2}[\sum_{k=p+q+1}^{M} \frac{[\mathrm{tr}(\widehat{\mathbf{S}}_k)]^2}{K^2} + \sum_{k=M+1}^{n-(P+1)} \frac{[\mathrm{tr}(\widehat{\mathbf{R}}_k)]^2}{K^2}] \; ; \quad (5.41)$$

and

$$KST = \sup_{1\leq j\leq n} |\widehat{Z}^m_{n,M}(j/n)| \; , \quad (5.42)$$

can be compared to the standard criteria of Hosking (1980) in (1.22),

$$\widehat{Q}_H^m = n\sum_{k=1}^{M}\mathrm{tr}(\widehat{\mathbf{C}}_k'\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{C}}_k\widehat{\mathbf{\Sigma}}^{-1}) \;;$$

and Li and McLeod (1981) in (1.24),

$$\widehat{Q}_{LM}^m = n\sum_{k=1}^{M}\mathrm{tr}(\widehat{\mathbf{C}}_k'\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{C}}_k\widehat{\mathbf{\Sigma}}^{-1}) + \frac{m^2M(M+1)}{2n} \;.$$

For a nominal level $\alpha$, the rejection regions for $CVT$ and $KST$ are based on the critical points of the corresponding functionals of the Brownian bridge. Regions associated to $\widehat{Q}_H^m$ and $\widehat{Q}_{LM}^m$ use the chi−square quantile $\chi_{m^2[M-(p+q)],\alpha}$. Comparisons are performed now, both in size and power, using simulation techniques.

### 5.5.1 Size

In principle, in $\widehat{Q}_H^m$ and $\widehat{Q}_{LM}^m$ the value of $M$ is taken of the order $O(\sqrt{n})$. It is however of interest to study the dependence on $M$ of the size of the four methods above. Two models are considered. First, a *VAR(1)* $\mathbf{X}_t = \mathbf{\Phi}_1\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$ in which

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.5603 & 0.5361 \\ -0.0366 & 0.6303 \end{pmatrix} \;. \tag{5.43}$$

The eigenvalues of the matrix $\mathbf{\Phi}_1$ in (5.43) are $\delta_j = 0.5953 \pm 0.1356\,i$, $j = 1, 2$, so that $|\delta_1| = |\delta_2| = 0.6106 < 1$. The associated eigenvectors are $\boldsymbol{\gamma}_1 = (0.9675, 0.0632 + 0.2447\,i)'$; $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (0.9675, 0.0632 - 0.2447\,i)'$. An additional *VMA(1)* process of the form $\mathbf{X}_t = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1\boldsymbol{\varepsilon}_{t-1}$ is analyzed, with

$$\mathbf{\Theta}_1 = \begin{pmatrix} 0.0589 & 0.3047 \\ -0.0093 & 0.1212 \end{pmatrix} \;. \tag{5.44}$$

The eigenvalues of $\mathbf{\Theta}_1$ in (5.44) are $\delta_j = 0.0901 \pm 0.0433\,i$, $j = 1, 2$, so that $|\delta_1| = |\delta_2| = 0.0999 < 1$. The eigenvectors are $\boldsymbol{\gamma}_1 = (0.9850, 0.1007 + 0.1400\,i)'$; $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (0.9850, 0.1007 - 0.1400\,i)'$. In both cases, the covariance matrix $\mathbf{\Sigma}$ of the errors is taken as in (5.30). For each model, $N = 1000$ independent replicas are generated. The sample size considered for the *VAR(1)* model is $n = 250$; and $n = 200$ for the *VMA(1)*. In both cases, the nominal level is $\alpha = .05$.

For values of $2 \leq M \leq 40$, figures 5.8 and 5.9 display the resulting empirical sizes for the *VAR(1)* and *VMA(1)* cases, respectively. Both plots indicate that the size of both $CVT$ in (5.41) and $KST$ in (5.42) is relatively stable with respect the value of

$M$. As it can be seen in these figures, this typically falls inside the horizontal bands $.05 \pm 1.96\sqrt{0.05 \times 0.95/1000}$. In contrast, the size of Hosking (1980) decreases. In turn, that of Li and McLeod (1981) is much less stable, being above the nominal level $\alpha = 0.05$ in the *VMA(1)* setting.

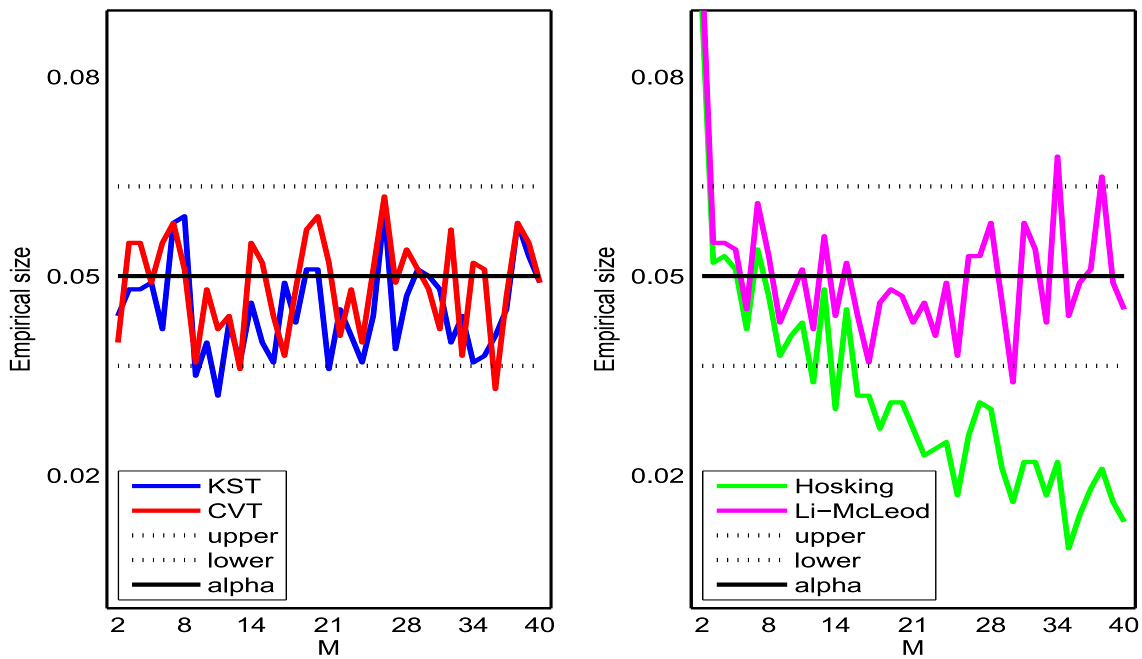


Figure 5.8: Comparison of empirical sizes for different values of the lag M for the *VAR(1)* model of section 5.5.1

### 5.5.2 Power

For making comparisons in power, two models are considered. First, the parametric *VARMA(1,1)* process $\mathbf{X}_t - \mathbf{\Phi}_1\mathbf{X}_{t-1} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1(\beta)\boldsymbol{\varepsilon}_{t-1}$ defined by the setting (5.36)−(5.37)−(5.30) of section 5.4.2. A *VAR(1)* is tested at the nominal level $\alpha = .05$. For each value of $0 \leq \beta < 1$, $N = 1000$ independent data samples of length $n = 200$ are generated. The value of the lag $M$ is taken as integer part of $\sqrt{n}$. Hence, $M = 14$. Results are displayed at the left part of figure 5.10.

Alternatively, a *VARMA(1,1)* $\mathbf{X}_t - \mathbf{\Phi}_1(\beta)\mathbf{X}_{t-1} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1\boldsymbol{\varepsilon}_{t-1}$ is considered, where $\mathbf{\Theta}_1$ is as in (5.29); $\mathbf{\Sigma}$ as in (5.30); and $\mathbf{\Phi}_1(\beta) = \beta\mathbf{\Phi}_1$, $0 \leq \beta < 1$, where

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.9177 & 0.8780 \\ -0.0599 & 1.0324 \end{pmatrix}. \tag{5.45}$$

The eigenvalues of $\mathbf{\Phi}_1$ in (5.45) are $\delta_j = 0.9750 \pm 0.2221\,i$, $j = 1, 2$, so that $|\delta_1| = |\delta_2| = 1.0000$. The eigenvectors are $\boldsymbol{\gamma}_1 = (0.9675, 0.0632 + 0.2447\,i)'$, and $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 =$

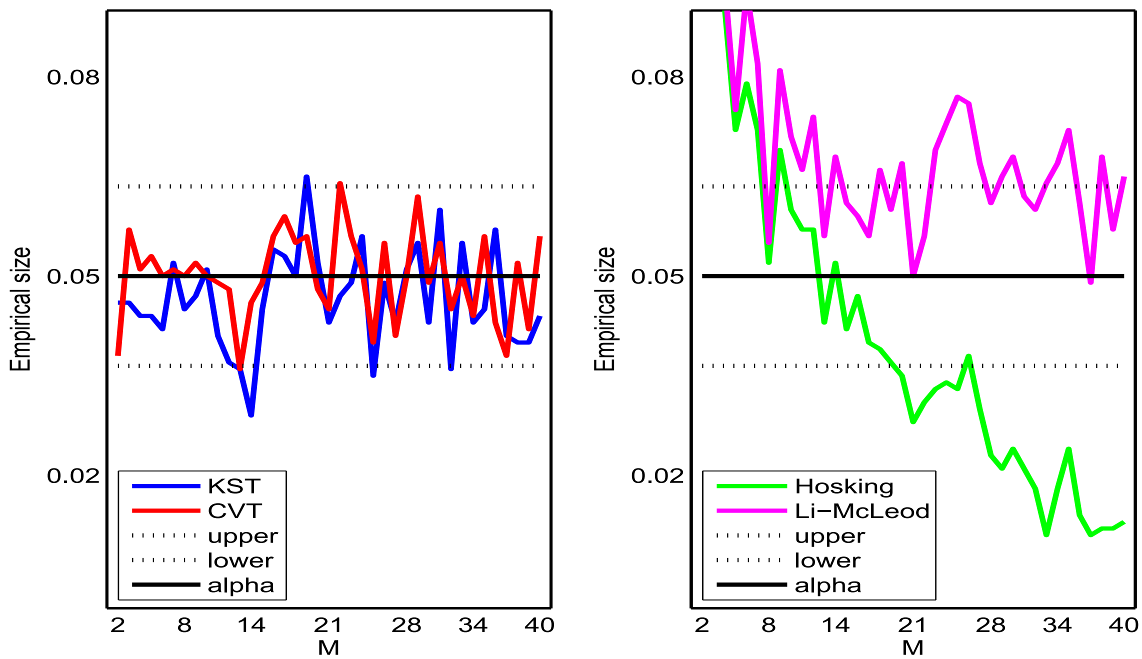


Figure 5.9: Comparison of empirical sizes for different values of the lag M for the *VMA(1)* model of section 5.5.1

$(0.9675, 0.0632 - 0.2447\,i)'$. A *VMA(1)* is now tested at level $\alpha = .05$. Choices for $N$, $n$, and $M$ are as before. Results are given at the right part of figure 5.10.

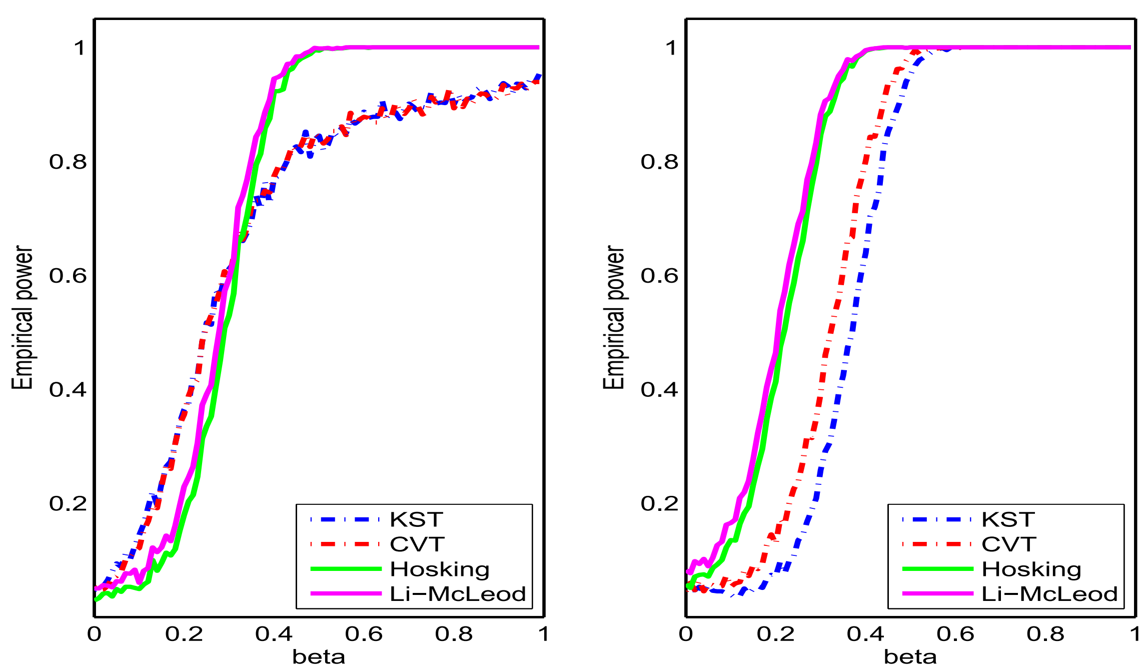


Figure 5.10: Comparison of empirical powers for a fixed value of the lag M in the first two simulation experiments of section 5.5.2. Left: *VAR(1)* fitted under a parametric *VARMA(1,1)* model; Right: *VMA(1)* fitted under a parametric *VARMA(1,1)* model

According to figure 5.10, our methods are locally more powerful than those by Hosking (1980) and Li and McLeod (1981) when a *VAR(1)* is tested. In turn, the latter procedures clearly outperform $KST$ and $CVT$ when the postulated null model is a *VMA(1)*. In this case, all the power functions tend to 1 when $\beta \to 1$.

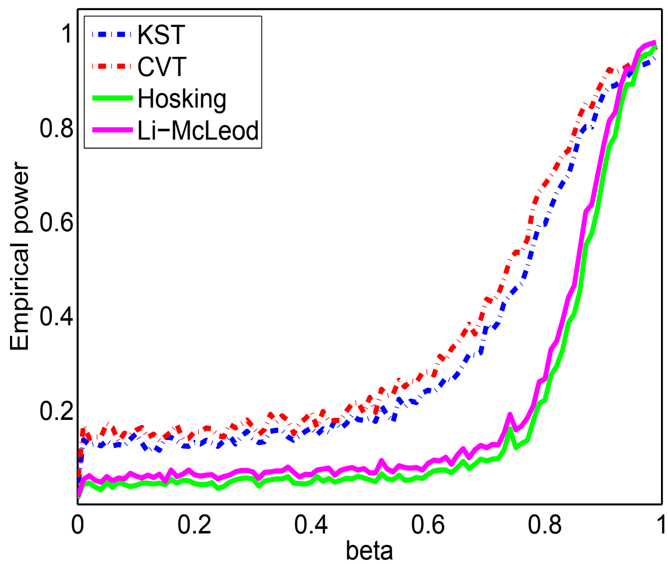


Figure 5.11: Comparison of empirical powers for a fixed value of the lag M in the third simulation experiment of section 5.5.2. A *VAR(2)* is fitted under a parametric *VARMA(2,2)* model

Finally, the parametric *VARMA(2,2)* model $\mathbf{X}_t - \mathbf{\Phi}_1\mathbf{X}_{t-1} - \mathbf{\Phi}_2\mathbf{X}_{t-2} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1(\beta)\boldsymbol{\varepsilon}_{t-1} + \mathbf{\Theta}_2(\beta)\boldsymbol{\varepsilon}_{t-2}$, $0 \leq \beta < 1$, of section 5.4.2 is revisited. A *VAR(2)* is now postulated and fitted. Choices for the tuning constants $N$, $n$, $M$, and $\alpha$ are as above. Results are in figure 5.11. As seen there, the empirical power of our procedures is above those of the standard methods by Hosking (1980) and Li and McLeod (1981).

## 5.6 A multivariate version of the cumulative periodogram statistic

This section explores a possible extension to the multivariate case of the usual univariate cumulative periodogram statistic. The basic idea is to use the analogy that exists for $m = 1$ between the adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ of (1.29), and the residual autocorrelations $\widehat{r}_k$ of (1.5).

Consider the discrete Fourier transform of the $m \times 1$ normalized residual vectors

$$\widehat{\mathbf{A}}(\omega) = \frac{1}{\sqrt{2\pi n}} \sum_{t>P}^{n} \exp(-i\omega t)\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\boldsymbol{\varepsilon}}_t \ , \quad 0 \leq \omega \leq \pi \ . \tag{5.46}$$

Following Brockwell and Davis (1991, chapter 11), the associated $m \times m$ standardized sample spectral density matrix is of the form

$$\widehat{\mathbf{I}}_n(\omega) = \widehat{\mathbf{A}}(\omega)\widehat{\mathbf{A}}^*(\omega) =$$

$$= \frac{1}{2\pi}\,\widehat{\mathbf{\Sigma}}^{-1/2}[\widehat{\mathbf{\Sigma}} + \sum_{k=1}^{n-(P+1)} \widehat{\mathbf{C}}_k \exp(ik\omega) + \sum_{k=1}^{n-(P+1)} \widehat{\mathbf{C}}'_k \exp(-ik\omega)]\widehat{\mathbf{\Sigma}}^{-1/2} \ , \quad 0 \le \omega \le \pi \ , \tag{5.47}$$

where the $m \times m$ matrices

$$\widehat{\mathbf{C}}_k = \frac{1}{n}\sum_{t>P}^{n-k} \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}'_{t+k} \ , \quad 0 \le k \le n-(P+1) \ ,$$

are as given in (1.20) (Chitturi, 1974), and $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{C}}_0$.

Recall now the definition of the $m \times m$ residual autocorrelation matrices of (1.21),

$$\widehat{\mathbf{R}}_k = \widehat{\mathbf{C}}'_k \widehat{\mathbf{\Sigma}}^{-1} \ , \quad 1 \le k \le n-(P+1) \ .$$

The trace of (5.47) is the squared Euclidean norm of $\widehat{\mathbf{A}}(\omega)$ in (5.46). Thus, after dividing by $\sqrt{m}$, the function

$$\widehat{I}_n^m(\omega) = \frac{1}{\sqrt{m}}\,\mathrm{tr}\,[\widehat{\mathbf{I}}_n(\omega)] =$$

$$= \frac{1}{2\pi}\,[\sqrt{m} + 2\sum_{k=1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}} \cos(k\omega)\,] \ , \quad 0 \le \omega \le \pi \ , \tag{5.48}$$

can be taken as a multivariate version of the standardized univariate residual periodogram (Ubierna and Velilla, 2007, sec. 1),

$$\widehat{I}_n(\omega) = \frac{1}{2\pi}\,[1 + 2\sum_{k=1}^{n-(P+1)} \widehat{r}_k \cos(k\omega)\,] \ , \quad 0 \le \omega \le \pi \ .$$

This is because, as seen before, for $m = 1$ the adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ of (1.29) coincide with the residual autocorrelations $\widehat{r}_k$ of (1.5).

Define the integrated version of the periodogram $\widehat{I}_n^m(\omega)$ in (5.48),

$$\widehat{F}_n^m(\pi u) = 2\int_0^{\pi u} \widehat{I}_n^m(\omega)d\,\omega =$$

$$= \frac{1}{\pi}\,[\sqrt{m}\,\pi u + 2\sum_{k=1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}} \frac{\sin(k\pi u)}{k}\,] \ , \quad 0 \le u \le 1 \ . \tag{5.49}$$

Consider also the $m \times m$ spectral density matrix of a $WN(\mathbf{0}, \mathbf{I}_m)$ sequence, $\mathbf{f}_0(\omega) = (1/2\pi)\,\mathbf{I}_m$, $-\pi \le \omega \le \pi$. Taking the trace and dividing by $\sqrt{m}$ leads to the univariate function $f_0^m(\omega) = \mathrm{tr}[\mathbf{f}_0(\omega)]/\sqrt{m} = \sqrt{m}/2\pi$, $-\pi \le \omega \le \pi$. Put $F_0^m(\pi u) = 2\int_0^{\pi u} f_0^m(\omega)d\,\omega = \sqrt{m}\,u$, $0 \le u \le 1$. The justification of definition (5.48) comes from the fact that the components of the residual process $\{\widehat{W}_n^m(u) : 0 \le u \le 1\}$ of (1.28),

$$\widehat{W}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k} \ ,$$

can be written in the form

$$\widehat{W}_n^m(u) = \sqrt{\frac{n}{2}}\,[\widehat{F}_n^m(\pi u) - F_0^m(\pi u)] = \sqrt{\frac{n}{2}}\,[\widehat{F}_n^m(\pi u) - \sqrt{m}\,u] \ . \tag{5.50}$$

The derivation of (5.50) follows closely the motivation given by Anderson (1993, sec. 1) of a general family of goodness-of-fit processes for spectral distributions. On the other hand, expression (5.50) leads to the construction of a multivariate version of the usual univariate cumulative periodogram statistic, as explained for example in Diggle (1990, p. 55) and Box et al. (1994, section 8.2.4).

Write $h = [n/2]$ for the integer part of $n/2$. In applications, the $KS$ statistic

$$\sup_{0\le u\le 1} |\widehat{W}_n^m(u)| = \sqrt{n/2}\sup_{0\le u\le 1} |\widehat{F}_n^m(\pi u) - \sqrt{m}\,u| \tag{5.51}$$

can be approximated by evaluating the sup at points $u_j = j/h$, $j = 1, \ldots, h$; and replacing $\widehat{F}_n^m(\pi j/h) = 2\int_0^{\pi j/h} \widehat{I}_n^m(\omega)d\,\omega$ by the Riemann sum

$$\widehat{U}_j^m = (4\pi/n)\sum_{k=1}^{j} \widehat{I}_n^m(2\pi k/n) \ , \quad j = 1, \ldots, h \ . \tag{5.52}$$

This amounts to replace $\sup_{0\le u\le 1} |\widehat{W}_n^m(u)|$ in (5.51) by

$$\widehat{C}_n^m = \sqrt{hm}\sup_{1\le j\le h} |(\widehat{U}_j^m/\sqrt{m}) - (j/h)| \ . \tag{5.53}$$

A plot in the unit square of the pairs

$$(j/h, \widehat{U}_j^m/\sqrt{m}) \ , \quad j = 1, \ldots, h \ , \tag{5.54}$$

may be called the cumulative periodogram of the $m\times 1$ residual vectors $\widehat{\boldsymbol{\varepsilon}}_t$, $P < t \le n$. The value of $\widehat{C}_n^m$ in (5.53) can be assessed graphically, superimposing on the plot two parallel bands to the left and to the right of the line $y = x$ at a distance $(hm)^{-1/2}KS_\alpha$, where $KS_\alpha$ is the appropriate quantile of the distribution of $\sup_{0\le u\le 1}|B(u)|$. This procedure will give only approximate significance levels for $\widehat{C}_n^m$. In fact, from theorem 3.4.1, the null asymptotic distribution of $\widehat{C}_n^m$ for $n$ is that of $\sup_{0\le u\le 1}|G^m(u)|$. The effect of this result in the size and power of $\widehat{C}_n^m$ has been illustrated in section 5.4.

The previous considerations suggest replacing the definition of $\widehat{I}_n^m(\omega)$ in (5.48) by the new set of modified residual periodogram ordinates

$$\widehat{J}_{n,M}^m(\omega) =$$

$$= \frac{1}{2\pi}\,[\sqrt{m} + 2\sum_{k=p+q+1}^{M}\frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}\cos(K\omega) + 2\sum_{k=M+1}^{n-(P+1)}\frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\cos(K\omega)\,]\ , \quad 0\le\omega\le\pi\ , \tag{5.55}$$

where $M\ge p+q+1$, and $K=k-(p+q)$. It is easy to verify that the components (5.40) of the truncated version $\{\widehat{Z}_{n,M}^m(u) : 0\le u\le 1\}$ of the modified process $\{\widehat{Z}_n^m(u) : 0\le u\le 1\}$ of (1.30), can be written in the form

$$\widehat{Z}_{n,M}^m(u) = \sqrt{\frac{n}{2}}\,[2\int_0^{\pi u}\widehat{J}_n^m(\omega)d\,\omega - \sqrt{m}\,u]\ . \tag{5.56}$$

Proceeding as above, identity (5.56) suggests considering the modified statistic

$$\widehat{C}_{n,M}^m = \sqrt{hm}\sup_{1\le j\le h}|(\widehat{V}_{j,M}^m/\sqrt{m}) - (j/h)|\ , \tag{5.57}$$

as an approximation of $KST = \sup_{1\le j\le n}|\widehat{Z}_{n,M}^m(j/n)|$ in (5.42), where $\widehat{V}_{j,M}^m = (4\pi/n)$ $\sum_{k=1}^{j}\widehat{J}_{n,M}^m(2\pi k/n)$, $j=1,\ldots,h$. From Theorem 4.4.1, the limit distribution of $\widehat{C}_{n,M}^m$ in (5.57) is exactly that of $\sup_{0\le u\le 1}|B(u)|$. The significance of $\widehat{C}_{n,M}^m$ can be assessed with a plot of the pairs

$$(j/h, \widehat{V}_{j,M}^m/\sqrt{m})\ , \quad j=1,\ldots,h\ ; \tag{5.58}$$

with superimposed lines $y = x\pm(hm)^{-1/2}KS_\alpha$. Unlike their counterparts of (5.54), the points $(j/h, \widehat{V}_{j,M}^m/\sqrt{m})$ will not have necessarily a monotonically increasing pattern, because definition (5.55) does not guarantee the condition $\widehat{J}_{n,M}^m(\omega)\ge 0$.

An example may help to clarify the use of the new cumulative periodograms defined by the pairs of (5.54) and (5.58). A sample of $n = 200$ observations is simulated from the bivariate *VAR(2)* model $\mathbf{X}_t = \mathbf{\Phi}_1\mathbf{X}_{t-1} + \mathbf{\Phi}_2\mathbf{X}_{t-2} + \boldsymbol{\varepsilon}_t$ defined by the roots in table 5.7, and the invertible matrix

$$\mathbf{A} = \begin{pmatrix} 1.2 & 0.4 \\ 0.2 & 1.0 \end{pmatrix}\ .$$

This leads to the matrices already considered in (5.31),

$$\mathbf{\Phi}_1 = \begin{pmatrix} -0.4964 & -0.0218 \\ 0.0091 & -0.5544 \end{pmatrix} \quad , \quad \mathbf{\Phi}_2 = \begin{pmatrix} 0.1286 & -0.0213 \\ 0.0089 & 0.0717 \end{pmatrix} .$$

The covariance matrix $\mathbf{\Sigma}$ of the errors is as in (5.30). A nominal level $\alpha = .05$ is used, so that $KS, .05 = 1.3582$. The value of $M$ is taken as $14 \cong \sqrt{n}$. A *VAR(1)* is postulated. For this generated sample, the approximate $KS$ statistics of (5.53) and (5.57) take the values $\widehat{C}_n^m = 1.2063$ and $\widehat{C}_{n,M}^m = 1.8292$, respectively. Hence, the lack of fit is only detected by the modified residual process of (5.40).

Figure 5.12 is the plot of the cumulative periodograms of (5.54) and (5.58). As seen there, the scatter of the $(j/h, \widehat{U}_j^m/\sqrt{m})$ lies always inside the superimposed bands $y = x \pm (hm)^{-1/2} KS_{.05}$, where $(hm)^{-1/2} KS_{.05} = 200^{-1/2} 1.3582 = 0.0960$. In turn, some of the modified $(j/h, \widehat{V}_{j,M}^m/\sqrt{m})$ are clearly outside this perimeter.

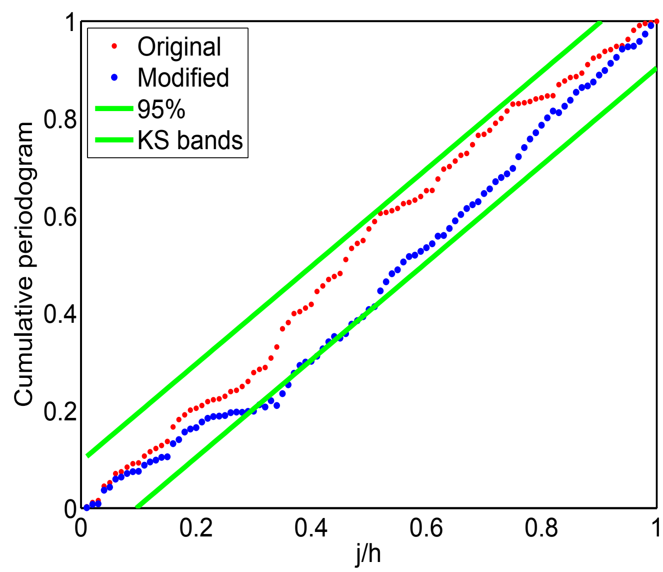


Figure 5.12: Multivariate original (red dots) and modified (blue dots) cumulative periodograms obtained when fitting a *VAR(1)* to a simulated sample of size $n = 200$ from the *VAR(2)* process given by the roots of table 5.7 and covariance matrix (5.30). The continuous lines are the 95% $KS$ confidence bands

## 5.7 A real data application

We finally illustrate the application of the tools of this chapter to a well-known real set of data. These refer to $n = 92$ quarterly, seasonally adjusted fixed investment $(Y_{t,1})$, disposable income $(Y_{t,2})$, and consumption expenditures $(Y_{t,3})$, in billions of DM, for the period between 1960 and 1982 in West Germany. They can be found in the supplementary `File E1` of the book by Lütkepohl (2005, section 3.2.3). The original data have a trend, that is removed by taking first differences of the logarithms. Hence, the goal is to model the $m = 3$ series $\mathbf{X}_t = (X_{t,1}, X_{t,2}, X_{t,3})'$, where

$$X_{t,j} = \log Y_{t+1,j} - \log Y_{t,j} \ , \quad j = 1, 2, 3 \ .$$

From Lütkepohl (2005, section 4.3.1) and Mahdi and McLeod (2012, sec. 4), a *VAR(2)* is considered adequate. We will study model selection on this data set by considering the fit of a nested sequence of *VAR(p)* models for $p$ = 1, 2, 3. A plot of both the original data and the first differences of the logarithms is given in figure 5.13.

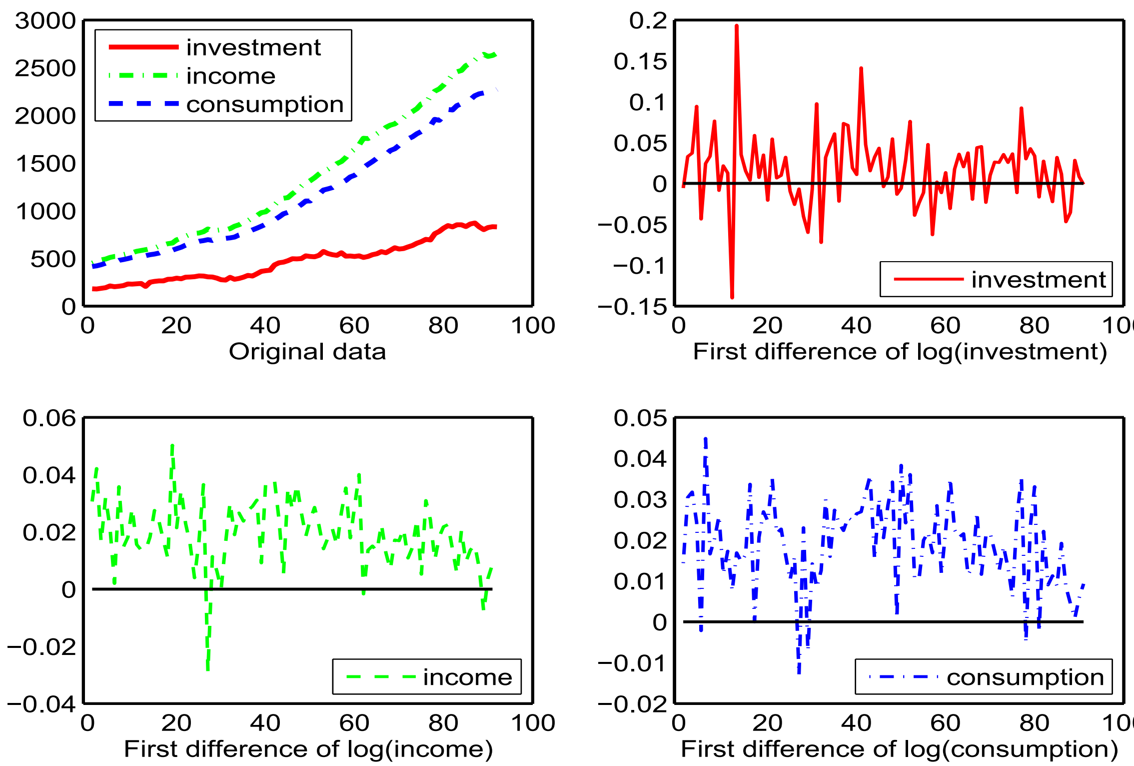


Figure 5.13: West German investment, income, and consumption data from 1960-1982: original observations, and first differences of logarithms

The estimated sample mean vector of the first differences of the logarithms of the investment, income, and consumption data is $(\overline{x}_1, \overline{x}_2, \overline{x}_3)' = (0.0168, 0.0195, 0.0187)'$. Table 5.16 reports the values of the ML estimates of the error covariance matrices and their corresponding determinants for *VAR(1)*, *VAR(2)*, and *VAR(3)* models. Table 5.17 displays the estimated autoregressive parameter matrices when fitting *VAR(p)* models to the West German data, $p$ = 1, 2, 3.

For a *VAR(1)* model, the eigenvalues associated with the estimated parameter matrix $\widehat{\mathbf{\Phi}}_1$ are $\delta_1 = 0.2226$, and $\delta_j = -0.3298 \pm 0.0391\, i$, $j$ = 2, 3. These satisfy the

| *VAR(p)* | $\widehat{\boldsymbol{\Sigma}}(p) \times 10^4$ | $\lvert\widehat{\boldsymbol{\Sigma}}(p)\rvert \times 10^{11}$ |
|---|---|---|
| $p = 1$ | $\begin{pmatrix} 18.3348 & 0.5182 & 1.3215 \\ 0.5182 & 1.2865 & 0.6743 \\ 1.3215 & 0.6743 & 1.0841 \end{pmatrix}$ | 1.5620 |
| $p = 2$ | $\begin{pmatrix} 17.6799 & 0.5548 & 1.2679 \\ 0.5548 & 1.1375 & 0.5761 \\ 1.2679 & 0.5761 & 0.8901 \end{pmatrix}$ | 1.0740 |
| $p = 3$ | $\begin{pmatrix} 17.5518 & 0.4938 & 1.1959 \\ 0.4938 & 1.0989 & 0.5328 \\ 1.1959 & 0.5328 & 0.8127 \end{pmatrix}$ | 0.9552 |

Table 5.16: ML estimates of the error covariance matrices for the West German data

| *VAR(p)* | $\widehat{\boldsymbol{\Phi}}_1$ | $\widehat{\boldsymbol{\Phi}}_2$ | $\widehat{\boldsymbol{\Phi}}_3$ |
|---|---|---|---|
| $p = 1$ | $\begin{pmatrix} -0.22 & 0.42 & 0.58 \\ 0.03 & -0.01 & 0.24 \\ 0.00 & 0.31 & -0.21 \end{pmatrix}$ | – | – |
| $p = 2$ | $\begin{pmatrix} -0.27 & 0.34 & 0.65 \\ 0.04 & -0.12 & 0.30 \\ 0.00 & 0.29 & -0.28 \end{pmatrix}$ | $\begin{pmatrix} -0.13 & 0.18 & 0.60 \\ 0.06 & 0.02 & 0.05 \\ 0.05 & 0.37 & -0.12 \end{pmatrix}$ | – |
| $p = 3$ | $\begin{pmatrix} -0.26 & 0.37 & 0.46 \\ 0.05 & -0.08 & 0.19 \\ 0.00 & 0.30 & -0.39 \end{pmatrix}$ | $\begin{pmatrix} -0.13 & 0.25 & 0.44 \\ 0.06 & 0.06 & -0.04 \\ 0.04 & 0.34 & -0.13 \end{pmatrix}$ | $\begin{pmatrix} 0.05 & 0.36 & -0.20 \\ 0.02 & 0.20 & -0.05 \\ 0.02 & 0.18 & 0.10 \end{pmatrix}$ |

Table 5.17: Yule-Walker estimates of the parameter matrices for the West German data

assumptions required in model (1.1), because $|\delta_1| = 0.2226 < 1$ and $|\delta_2| = |\delta_3| = 0.3321 < 1$. If a *VAR(2)* model and a *VAR(3)* model are estimated, the $mp$ roots of the determinantal equation $|\widehat{\boldsymbol{\Phi}}(z)| = 0$, where $m = 3$, are given in Tables 5.18 and 5.19, respectively. For each of these models, the roots have been computed using the method of Appendix 5.2. They are all different from each other, and they all lie outside the unit circle.

As in Lütkepohl (2005, section 4.3.1), the statistic

$$AIC = \log|\widehat{\boldsymbol{\Sigma}}| + \frac{2}{n} p\, m^2$$

| $j$ | $\varsigma_j$ | $\lvert\varsigma_j\rvert$ |
|---|---|---|
| 1 | $1.6657 + 0.0000\,i$ | 1.6657 |
| 2 | $-1.0994 + 1.4235\,i$ | 1.7986 |
| 3 | $-1.0994 - 1.4235\,i$ | 1.7986 |
| 4 | $-0.4419 - 1.9929\,i$ | 2.0413 |
| 5 | $-0.4419 + 1.9929\,i$ | 2.0413 |
| 6 | $-2.5738 + 0.0000\,i$ | 2.5738 |

Table 5.18: Roots of the determinantal equation $|\boldsymbol{\Phi}(z)| = 0$ when fitting a trivariate *VAR(2)* model to the West German data

| $j$ | $\varsigma_j$ | $\lvert\varsigma_j\rvert$ |
|---|---|---|
| 1 | $15.1130 + 0.0000\,i$ | 15.1130 |
| 2 | $2.8308 + 0.0000\,i$ | 2.8308 |
| 3 | $-1.4788 + 1.1933\,i$ | 1.9003 |
| 4 | $-1.4788 - 1.1933\,i$ | 1.9003 |
| 5 | $1.3484 + 0.0000\,i$ | 1.3484 |
| 6 | $-0.6594 + 1.8464\,i$ | 1.9606 |
| 7 | $-0.6594 - 1.8464\,i$ | 1.9606 |
| 8 | $-0.7853 + 1.2367\,i$ | 1.4650 |
| 9 | $-0.7853 - 1.2367\,i$ | 1.4650 |

Table 5.19: Roots of the determinantal equation $|\boldsymbol{\Phi}(z)| = 0$ when fitting a trivariate *VAR(3)* model to the West German data

selects a *VAR(2)* process, where $m = 3$. Results for the standard criteria of Hosking (1980) in (1.22); and of Li and McLeod (1981) in (1.24), are presented in Table 5.20. The P-values associated with these two methods support typically a *VAR(1)* model, and a *VAR(2)* model only for low values of the lag $M$.

Table 5.21 displays the behavior of the goodness-of-fit statistics analyzed previously, including those computed for the residual process $\{\widehat{W}_n^m(u) : 0 \le u \le 1\}$ of (1.28); the fully modified process $\{\widehat{Z}_n^m(u) : 0 \le u \le 1\}$ of (1.30); and the truncated modified process of (5.40). The results of Table 5.21 indicate just moderate evidence against a *VAR(1)* model, confirming in general the *VAR(2)* specification. In fact, some of the Cramér-von Mises statistics point towards a *VAR(3)*.

The techniques of this chapter are of asymptotic nature, and a sample size of $n = 92$ for $m = 3$ is perhaps not large enough for extracting definitive conclusions. This is because the number of autoregressive parameters to be estimated range from $m^2 = 9$, for $p = 1$, and $3m^2 = 27$, for $p = 3$. However, it seems that, independently of the choice of the lag $M$, table 5.21 offers some additional evidence for supporting the choice of Lütkepohl (2005, section 4.3.1) of a *VAR(2)* for this data set.

| | *VAR(1)* | *VAR(2)* | *VAR(3)* |
|---|---|---|---|
| AIC | -24.6847 | -24.8614$^\star$ | -24.7808 |
| | | HOSKING | |
| **Statistics** | | | |
| $M=5$ | 50.0695 | 29.7402 | 24.5617 |
| $M=10$ | 89.6045 | 68.1544 | 63.2782 |
| $M=15$ | 131.9975 | 112.0166 | 107.1634 |
| **P-values** | | | |
| $M=5$ | 0.0597 | 0.3260$^\star$ | 0.1375 |
| $M=10$ | 0.2402 | 0.6067 | 0.4665 |
| $M=15$ | 0.3395 | 0.6129 | 0.5047 |
| | | LI-MCLEOD | |
| **Statistics** | | | |
| $M=5$ | 51.5530 | 31.2237 | 26.0452 |
| $M=10$ | 95.0441 | 73.5940 | 68.7178 |
| $M=15$ | 143.8656 | 123.8848 | 119.0315 |
| **P-values** | | | |
| $M=5$ | 0.0449 | 0.2621$^\star$ | 0.0987 |
| $M=10$ | 0.1362 | 0.4257 | 0.2899 |
| $M=15$ | 0.1318 | 0.3138 | 0.2201 |

Table 5.20: Comparisons between goodness-of-fit statistics and P-values for the West German data.

For *VAR(1)*, *VAR(2)*, and *VAR(3)* models, the plots of the adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ of (1.29), and of the modified $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ of (4.11), are given in figures 5.14, 5.15, and 5.16, respectively. The dashed lines are the estimated 95% $KS$ confidence bands of (2.41).

$$\pm z_{.025}\, n^{-1/2}\sqrt{1-\mathbf{a}'_m \Xi'_k (\mathbf{Z}'_M \mathcal{W}^{-1}\mathbf{Z}_M)^{-1}\Xi_k \mathbf{a}_m}\ , \quad 1 \le k \le M\ ,$$

where $z_{.025} = 1.96$, $n = 92$, and $M = 9 = [\sqrt{92}]$. The continuous lines are the 95% $KS$ pivotal confidence bands $\pm z_{.025}\, n^{-1/2}$ for the modified adjusted traces. As expected from (4.32) in Chapter 4, in all the plots the modified traces $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ can be approximated by the adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ after the first few lags. Figures 5.14, 5.15, and 5.16 indicate only a discrepancy for the *VAR(1)* case, thus offering again some graphical evidence in favor of a *VAR(2)* process.

| | *VAR(1)* | *VAR(2)* | *VAR(3)* |
|---|---|---|---|
| | | ORIGINAL RESIDUAL PROCESS | |
| KSRG | 1.2318 | 0.5454 | 0.3760 |
| **P-value** | 0.0970 | 0.9124 | 0.9987 |
| CVRG | 0.3231 | 0.0576 | 0.0294 |
| **P-value** | 0.1200 | 0.8300 | 0.9800 |
| | | MODIFIED PROCESS | |
| KSM | 1.1708 | 1.0313 | 0.7680 |
| **P-value** | 0.1294 | 0.2392 | 0.5936 |
| CVM | 0.3627 | 0.3829 | 0.1836 |
| **P-value** | 0.0900 | 0.0800 | 0.3000 |
| | | TRUNCATED MODIFIED PROCESS | |
| KST | | | |
| $M = 5$ | 1.1709 | 1.0307 | 0.7586 |
| $M = 10$ | 1.1708 | 1.0313 | 0.7680 |
| $M = 15$ | 1.1708 | 1.0313 | 0.7681 |
| **P-value** | | | |
| $M = 5$ | 0.1294 | 0.2392 | 0.6104 |
| $M = 10$ | 0.1294 | 0.2392 | 0.5936 |
| $M = 15$ | 0.1294 | 0.2392 | 0.5936 |
| CVT | | | |
| $M = 5$ | 0.3627 | 0.3828 | 0.1845 |
| $M = 10$ | 0.3627 | 0.3829 | 0.1836 |
| $M = 15$ | 0.3627 | 0.3829 | 0.1836 |
| **P-value** | | | |
| $M = 5$ | 0.0900 | 0.0800 | 0.3000 |
| $M = 10$ | 0.0900 | 0.0800 | 0.3000 |
| $M = 15$ | 0.0900 | 0.0800 | 0.3000 |
| | | KS CUMULATIVE | |
| RESIDUAL | 1.1724 | 0.5902 | 0.4844 |
| MODIFIED | 1.1497 | 1.0959 | 0.8289 |

Table 5.21: Comparisons between functionals of the goodness-of-fit processes for the West German data.

Figure 5.17 is the plot of the multivariate original cumulative periodogram of (5.54), $(j/h, \widehat{U}_j^m/\sqrt{m})$, obtained when fitting a *VAR(1)*, a *VAR(2)* and a *VAR(3)* to the West German data. Figure 5.18 displays the multivariate modified cumulative periodograms $(j/h, \widehat{V}_{j,M}^m/\sqrt{m})$ of (5.58). The bands

$$y = x \pm (hm)^{-1/2} KS_{.05} \ ,$$

where $h = 45 = [n/2] - 1$ and $(hm)^{-1/2} KS_{.05} = (45 \times 3)^{-1/2} 1.3582 = 0.1169$, are also included. As before, some graphical indication against a *VAR(1)* process, and in favor of a *VAR(2)* specification, is found.

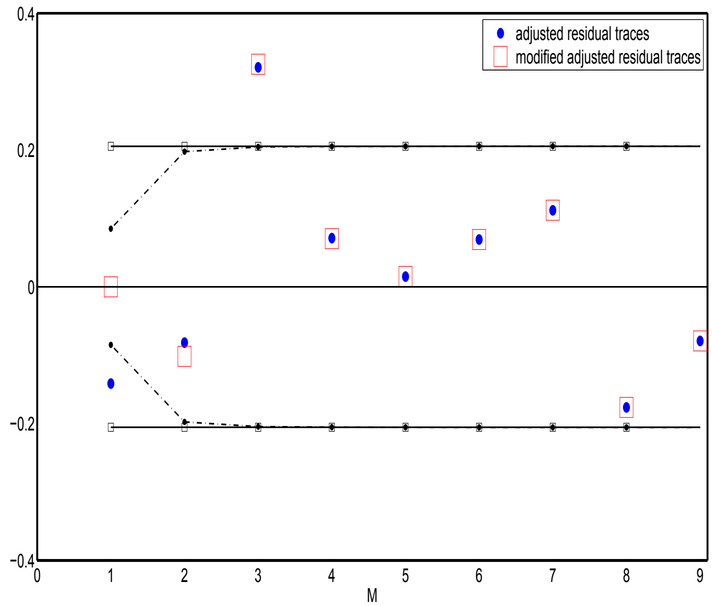


Figure 5.14: Scatter plot of the adjusted residual traces and the modified traces when fitting a *VAR(1)* model to the West German data. The dashed lines are the 95% $KS$ confidence bands for the adjusted residual traces. The continuous lines are the 95% $KS$ confidence bands for the modified traces.

## 5.8 Summary and conclusions

This thesis proposes a new goodness-of-fit process for *VARMA(p,q)* models. Our results generalize those obtained by Ubierna and Velilla (2007). Some preliminary findings are derived first. In particular, we derive in the multivariate case an explicit form of the information matrix as a limit. A simplified version of the basic highly dimensional relation between residual and error sample covariance matrices in Hosking (1980) is suggested. This leads to the introduction of the adjusted residual traces, whose properties are investigated in some detail. This new relation is similar in structure to that considered by Box and Pierce (1970) in the univariate case.

The starting point is to extend to a multivariate setting the goodness-of-fit processes studied for *ARMA(p,q)* models by Durlauf (1991) and Anderson (1993). We derive weak convergence of this extension, that is based on the unfeasible error vectors, as well as for its residual version. An explicit form of the limit covariance function is obtained. Since this depends on unknown parameters, directly assessing

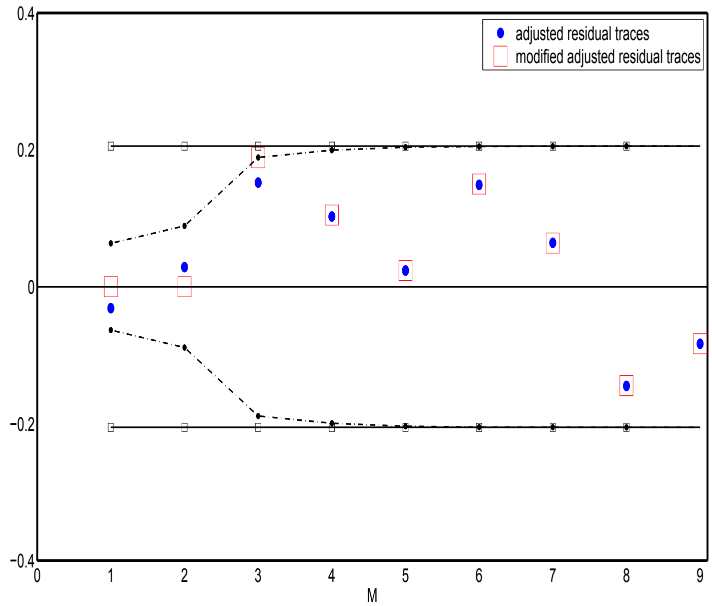


Figure 5.15: Scatter plot of the adjusted residual traces and the modified traces when fitting a *VAR(2)* model to the West German data. The dashed lines are the 95% $KS$ confidence bands for the adjusted residual traces. The continuous lines are the 95% $KS$ confidence bands for the modified traces.

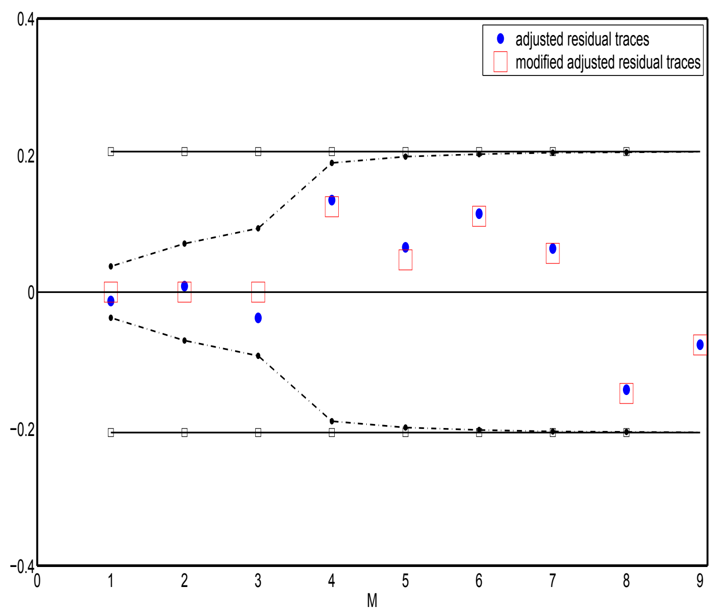


Figure 5.16: Scatter plot of the adjusted residual traces and the modified traces when fitting a *VAR(3)* model to the West German data. The dashed lines are the 95% $KS$ confidence bands for the adjusted residual traces. The continuous lines are the 95% $KS$ confidence bands for the modified traces.

significance for goodness-of-fit purposes is not feasible. As an alternative, we introduce a sequence of modified adjusted residual traces, that are used to construct a

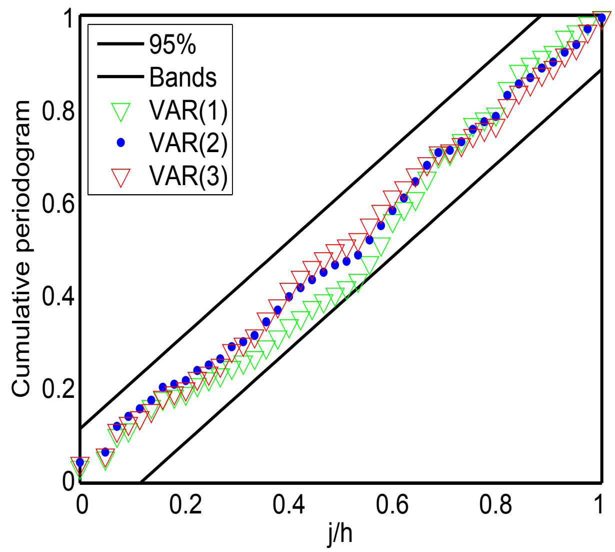


Figure 5.17: Multivariate original cumulative periodograms obtained when fitting a *VAR(1)* (green), a *VAR(2)* (blue) and a *VAR(3)* (red) to the West German data. The continuous lines are the 95% $KS$ confidence bands

.

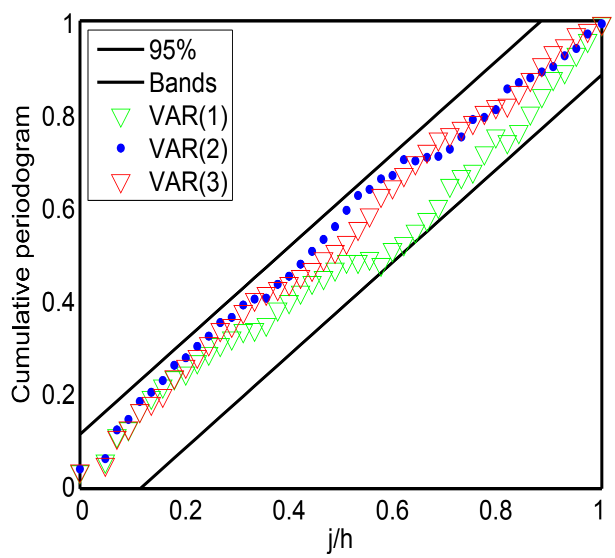


Figure 5.18: Multivariate modified cumulative periodograms obtained when fitting a *VAR(1)* (green), a *VAR(2)* (blue) and a *VAR(3)* (red) to the West German data. The continuous lines are the 95% $KS$ confidence bands

.

goodness-of-fit process that converges weakly to the Brownian bridge. Consequently, we can assess goodness-of-fit of a *VARMA(p,q)* model by using statistics defined by continuous functionals whose null pivotal distribution is tabulated. As for instance, the Kolmogorov-Smirnov statistic and the Cramér-von Mises statistic.

In this chapter, we study illustrative examples supporting the theoretical results given in Chapters 2, 3 and 4. Simulations and comparisons in size, power and dependence on the lag $M$ show that our methods are effective for detecting lack of fit. A multivariate version of the usual univariate cumulative periodogram statistic is also introduced. An application with a well-known real data set illustrates how to use these procedures for identifying a proper multivariate time series model.

## Appendix 5.1: Structure of the leading matrices $\mathbf{P}_{jj}$ in *VAR(p)* and *VMA(q)* models

For studying the *VAR(p)* case, write $\mathcal{W}^{-1/2}\mathbf{Z}_p = \mathcal{X}_p \otimes \mathbf{\Sigma}^{-1/2}$, where:

$$\mathcal{X}_p = \begin{pmatrix} \mathbf{\Sigma}^{1/2}\mathbf{H}_0' & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Sigma}^{1/2}\mathbf{H}_1' & \mathbf{\Sigma}^{1/2}\mathbf{H}_0' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Sigma}^{1/2}\mathbf{H}_2' & \mathbf{\Sigma}^{1/2}\mathbf{H}_1' & \mathbf{\Sigma}^{1/2}\mathbf{H}_0' & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}^{1/2}\mathbf{H}_{p-1}' & \mathbf{\Sigma}^{1/2}\mathbf{H}_{p-2}' & \mathbf{\Sigma}^{1/2}\mathbf{H}_{p-3}' & \cdots & \mathbf{\Sigma}^{1/2}\mathbf{H}_0' \end{pmatrix} = \begin{pmatrix} \mathbf{\Upsilon}_1' \\ \vdots \\ \mathbf{\Upsilon}_p' \end{pmatrix} ,$$

and the $m \times m$ matrices $\{\mathbf{H}_k : k \geq 0\}$ are the coefficients of the series expansion $\mathbf{\Phi}^{-1}(z) = \sum_{k=0}^{\infty} \mathbf{H}_k \mathbf{z}^k$. From expression (2.70) in appendix 2.1, the information matrix of a *VAR(p)* model is of the form

$$\mathbf{I}[\mathrm{vec}(\mathbf{\Phi})] = \mathbf{A} \otimes \mathbf{\Sigma}^{-1} ,$$

where $\mathbf{A}$ is a $pm \times pm$ matrix with $(r, R)$ block $\sum_{k=0}^{\infty} \mathbf{H}_{k-r}\mathbf{\Sigma}\mathbf{H}_{k-R}'$, $r, R = 1, \ldots, p$. Hence, (5.27) follows from

$$\mathbf{P}_{jj} = \mathbf{\Xi}_j'\mathbf{I}^{-1}[\mathrm{vec}(\mathbf{\Phi})]\mathbf{\Xi}_j = (\Upsilon_j' \otimes \mathbf{\Sigma}^{-1/2})(\mathbf{A}^{-1} \otimes \mathbf{\Sigma})(\Upsilon_j \otimes \mathbf{\Sigma}^{-1/2}) = \mathbf{A}_j \otimes \mathbf{I}_m ,$$

where $\mathbf{A}_j = \Upsilon_j'\mathbf{A}^{-1}\Upsilon_j \cong \Upsilon_j'(\mathcal{X}_p'\mathcal{X}_p)^{-1}\Upsilon_j = \mathbf{I}_m$, $j = 1, \ldots, p$.

On the other hand, according to the findings of section 2.3, the information matrix of a *VMA(q)* model can be written

$$\mathbf{I}[\mathrm{vec}(\mathbf{\Theta})] = \mathbf{\Sigma} \otimes \mathbf{B} ,$$

where $\mathbf{B}$ is a $qm \times qm$ matrix with $(s, S)$ block $\sum_{k=0}^{\infty} \mathbf{L}_{k-s}'\mathbf{\Sigma}^{-1}\mathbf{L}_{k-S}$, $s, S = 1, \ldots, q$. In the *VMA(q)* case, $\mathcal{W}^{-1/2}\mathbf{Z}_q = \mathbf{\Sigma}^{1/2} \otimes \mathcal{Y}_q$, where

$$\mathcal{Y}_p = \begin{pmatrix} \mathbf{\Sigma}^{-1/2}\mathbf{L}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Sigma}^{-1/2}\mathbf{L}_1 & \mathbf{\Sigma}^{-1/2}\mathbf{L}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Sigma}^{-1/2}\mathbf{L}_2 & \mathbf{\Sigma}^{-1/2}\mathbf{L}_1 & \mathbf{\Sigma}^{-1/2}\mathbf{L}_0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}^{-1/2}\mathbf{L}_{q-1} & \mathbf{\Sigma}^{-1/2}\mathbf{L}_{q-2} & \mathbf{\Sigma}^{-1/2}\mathbf{L}_{q-3} & \cdots & \mathbf{\Sigma}^{-1/2}\mathbf{L}_0 \end{pmatrix} = \begin{pmatrix} \mathbf{\Upsilon}_1' \\ \vdots \\ \mathbf{\Upsilon}_q' \end{pmatrix} ,$$

and the $m \times m$ matrices $\{\mathbf{L}_k : k \geq 0\}$ are the coefficients of the series expansion $\mathbf{\Theta}^{-1}(z) = \sum_{k=0}^{\infty} \mathbf{L}_k \mathbf{z}^k$ of (2.14). Thus:

$$\mathbf{P}_{jj} = \mathbf{\Xi}_j'\mathbf{I}^{-1}[\mathrm{vec}(\mathbf{\Theta})]\mathbf{\Xi}_j = (\mathbf{\Sigma}^{1/2} \otimes \Upsilon_j')(\mathbf{\Sigma}^{-1} \otimes \mathbf{B}^{-1})(\mathbf{\Sigma}^{1/2} \otimes \Upsilon_j) = \mathbf{I}_m \otimes \mathbf{B}_j , \quad (5.59)$$

where $\mathbf{B}_j = \Upsilon_j'\mathbf{B}^{-1}\Upsilon_j \cong \Upsilon_j'(\mathcal{Y}_q'\mathcal{Y}_q)^{-1}\Upsilon_j = \mathbf{I}_m$, $j = 1, \ldots, q$.

Expressions (5.27) and (5.29), that are dual from each other, confirm the findings in the *VAR(p)* and *VMA(q)* models presented in section 5.2.

## Appendix 5.2: Finding the roots of the determinantal equation $|\mathbf{\Phi}(z)| = 0$

Consider the determinantal equation

$$P(z) = |\mathbf{\Phi}(z)| = 0, \tag{5.60}$$

where $\mathbf{\Phi}(z) = \mathbf{I}_m - \mathbf{\Phi}_1 z - \cdots - \mathbf{\Phi}_p z^p$ depends on some known $m \times m$ real matrices $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p$. By construction,

$$P(z) = |\mathbf{\Phi}(z)| = 1 - \sum_{j=1}^{mp} a_j z^j \tag{5.61}$$

is a polynomial of degree $mp$. However, each of the entries of the $m \times m$ matrix $\mathbf{\Phi}(z) = (\Phi_{IJ}(z) : I, J : 1, \ldots, m)$ is a polynomial of degree $p$. Therefore, for large values of either $p$ or $m$, the exact values of the coefficients $\{a_j\}$ of the polynomial $P(z)$ in (5.61) are difficult to find. As a conclusion, $P(z)$ is a polynomial of unknown coefficients but of known value.

A possible method for determining the roots of the determinantal equation (5.60) is to proceed as follows:

**1.** Evaluate $P(z)$ at $mp + 1$ given and different points $z_0 = 0$, $z_1$, $z_2$, $\ldots$, $z_{mp}$.

**2.** Form the linear system

$$\begin{pmatrix} P(z_0) \\ P(z_1) \\ P(z_2) \\ \vdots \\ P(z_{mp}) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & z_1 & \cdots & z_1^{mp} \\ 1 & z_2 & \cdots & z_2^{mp} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{mp} & \cdots & z_{mp}^{mp} \end{pmatrix} \begin{pmatrix} 1 \\ -a_1 \\ -a_2 \\ \vdots \\ -a_{mp} \end{pmatrix} . \tag{5.62}$$

**3.** The linear system of (5.62) has a Vandermonde determinant

$$\prod_{0 \le i < j \le mp} (z_j - z_i) \ne 0 . \tag{5.63}$$

Therefore, the unknown coefficients $a_j$, $j = 1, \ldots, mp$, of (5.61) are the last $mp$ coordinates of the $(mp + 1) \times 1$ column vector

$$- \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & z_1 & \cdots & z_1^{mp} \\ 1 & z_2 & \cdots & z_2^{mp} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{mp} & \cdots & z_{mp}^{mp} \end{pmatrix}^{-1} \begin{pmatrix} P(z_0) \\ P(z_1) \\ P(z_2) \\ \vdots \\ P(z_{mp}) \end{pmatrix} . \tag{5.64}$$

Another possibility is to form

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{mp} \end{pmatrix} = - \begin{pmatrix} z_1 & \cdots & z_1^{mp} \\ z_2 & \cdots & z_2^{mp} \\ \vdots & \ddots & \vdots \\ z_{mp} & \cdots & z_{mp}^{mp} \end{pmatrix}^{-1} \begin{pmatrix} P(z_1) - 1 \\ P(z_2) - 1 \\ \vdots \\ P(z_{mp}) - 1 \end{pmatrix} . \tag{5.65}$$

The inverse in (5.65) exists, because the value of the determinant of the associated matrix at the right-hand side coincides with expression (5.63).

In applications, the magnitude $\prod_{0 \leq i < j \leq mp}(z_j - z_i)$ can take a large value, making this procedure numerically unstable. Thus, a convenient protocol for selecting the $z_j$ is to choose them sequentially, so that $z_0 = 0$ and for each $q = 1, \ldots, mp$, $\prod_{0 \leq i < j \leq q}(z_j - z_i) = 1$. Specifically:

**1.** For $q = 1$, $z_1$ must be the a root of the equation $x = 1$;

**2.** For $q = 2$, $z_2$ must satisfy $(x - z_1)x = 1$;

**3.** For $q = 3$, $z_3$ solves $(x - z_2)(x - z_1)x = 1$;

… …

According to the above, the $z_j$ can be easily found as roots of a sequence of recursive polynomials of degree $j$th each. In particular, $z_1 = 1$, and for $2 \leq j \leq mp$, $z_j$ solves

$$P_j(x) = [\prod_{k=1}^{j-1}(x - z_k)]x = 1 . \tag{5.66}$$

The coefficients of (5.66) can be easily found using the recursion $P_1(x) = x$,

$$P_{j+1}(x) = (x - z_j)P_j(x) , \quad j = 1, \ldots, mp - 1 . \tag{5.67}$$

# Chapter 6

# Further research

This thesis develops goodness-of-fit methods for assessing a causal and invertible $m-$variate autoregressive moving average *VARMA(p,q)* process of the form (1.1),

$$\mathbf{\Phi}(B)(\mathbf{X}_t - \boldsymbol{\mu}) = \mathbf{\Theta}(B)\boldsymbol{\varepsilon}_t \ .$$

A basic assumption is that true error vectors $\{\boldsymbol{\varepsilon}_t : t \in \mathbf{Z}\}$ behave as a zero mean white noise sequence $WN(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a $m \times m$ positive definite matrix. Therefore, a possible line for future research is to consider more flexible dependence structures for the $\{\boldsymbol{\varepsilon}_t : t \in \mathbf{Z}\}$. It is also interesting to consider a distribution of the error vectors different form the usual Gaussian model. This is a challenging and computationally demanding approach.

The methods proposed in this thesis are based on the adjusted residual traces of (1.29), that are given by

$$\frac{1}{\sqrt{m}} \begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{R}}_1) \\ \mathrm{tr}(\widehat{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{R}}_M) \end{pmatrix} = (\mathbf{I}_M \otimes \mathbf{a}'_m)\widehat{\mathbf{W}}^{-1/2} \begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{C}}'_1) \\ \mathrm{vec}(\widehat{\mathbf{C}}'_2) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{C}}'_M) \end{pmatrix} ,$$

where $\widehat{\mathbf{W}} = \mathbf{I}_M \otimes \widehat{\mathbf{C}}_0 \otimes \widehat{\mathbf{C}}_0$. As analyzed in Chapter 2, the large sample distribution of the random object above can be characterized, to some approximation, as

$$\sqrt{n}\,[\frac{1}{\sqrt{m}} \begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{R}}_1) \\ \mathrm{tr}(\widehat{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{R}}_M) \end{pmatrix}] \cong \mathbf{N}_M[\mathbf{0}, \mathbf{I}_M - (\mathbf{I}_M \otimes \mathbf{a}'_m)\mathbf{P}_M(\mathbf{I}_M \otimes \mathbf{a}_m)] \ .$$

The covariance matrix in the previous expression depends on the parameters of the model (1.1). Hence, a new set of modified residual traces

$$\frac{1}{\sqrt{m}} \begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{S}}_1) \\ \mathrm{tr}(\widehat{\mathbf{S}}_2) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{S}}_M) \end{pmatrix}$$

is suggested, that depend on a suitable sequence of transformation matrices $\{\mathbf{\Psi}_M\}$, that satisfy $\mathbf{\Psi}'_M\mathbf{\Psi}_M = \mathbf{I}_{m^2[M-(p+q)]}$ and $\mathbf{\Psi}'_M\mathcal{W}^{-1/2}\mathbf{Z}_M = \mathbf{0}$. The structure of the $\{\mathbf{\Psi}_M\}$, that make the distribution of the $\mathrm{tr}(\widehat{\mathbf{S}}_k)$ of pivotal nature, should be studied in more detail for numerical purposes.

Another line of extensions is to study the behavior of the technique in additional data sets. The moving average schemes are investigated using the general

*VARMA(p,q)* likelihood algorithm by Lütkepohl (2005). MLE estimates can be determined from simulated observations, but not in general from real data. This is a methodological limitation here, whose possible solution should be analyzed.

Finally, other econometric models could be considered for goodness-of-fit under the setting of this thesis. For example, multivariate volatility models, and multivariate GARCH models. Other dependence structures for the data, alternative to the *VARMA(p,q)* model of (1.1), could be also studied. As for instance, trends, structural break, seasonality, and cyclic variation.

# Bibliography

Anderson, T. W., 1955. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proceedings of the American Mathematical Society 6 (2), pp. 170–176.

Anderson, T. W., 1993. Goodness of fit tests for spectral distributions. The Annals of Statistics 21 (2), pp. 830–847.

Ash, R. B., Gardner, M. F., 1975. Topics in stochastic processes. Academic Press, New York.

Battaglia, F., 1990. Approximate power of portmanteau tests for time series. Statistics & Probability Letters 9 (4), 337 – 341.

Billingsley, P., 1976. Convergence of Probability Measures. Wiley, New York.

Bouhaddioui, C., Roy, R., 2006. A generalized portmanteau test for independence of two infinite-order vector autoregressive series. Journal of Time Series Analysis 27 (4), 505–544.

Box, G., Jenkins, G., Reinsel, G., 1994. Time Series Analysis: Forecasting and Control., 3rd Edition. Prentice-Hall, Englewood Cliffs, NJ.

Box, G. E. P., Pierce, D. A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American Statistical Association 65 (332), pp. 1509–1526.

Brockwell, P. J., Davis, R. A., 1991. Time series: Theory and methods. 2nd Edition, SpringerVerlag, New York.

Bruce, A. G., Martin, R. D., 1989. Leave-k-out diagnostics for time series. Journal of the Royal Statistical Society. Series B (Methodological) 51 (3), pp. 363–424.

Chen, W. W., Deo, R. S., 2004. A generalized portmanteau goodness-of-fit test for time series models. Econometric Theory 20 (2), 382–416.

Chitturi, R. V., 1974. Distribution of residual autocorrelations in multiple autoregressive schemes. Journal of the American Statistical Association 69 (348), 928–934.

Chitturi, R. V., 1976. Distribution of multivariate white noise autocorrelations. Journal of the American Statistical Association 71 (353), pp. 223–226.

Davies, N., Triggs, C. M., Newbold, P., 1977. Significance levels of the box-pierce portmanteau statistic in finite samples. Biometrika 64 (3), pp. 517–522.

Delgado, M. A., Velasco, C., 2010. Distribution-free tests for time series models specification. Journal of Econometrics 155 (2), 128 – 137.

Diggle, P., 1990. Time Series: A Biostatistical Introduction. Oxford University Press, Oxford.

Dunsmuir, W., Hannan, E. J., 1976. Vector linear time series models. Advances in Applied Probability 8 (2), pp. 339–364.

Durbin, J., 1975. Tests of model specification based on residuals. Vol. A Survey of Statistical Design and Linear Models. North-Holland, Amsterdam.

Durlauf, S. N., 1991. Spectral based testing of the martingale hypothesis. Journal of Econometrics 50 (3), 355 – 376.

Fang, K.-T., Zhang, Y.-T., 1990. Generalized Multivariate Analysis. Springer-Verlag.

Francq, C., Rassi, H., 2007. Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors. Journal of Time Series Analysis 28 (3), 454 – 470.

Francq, C., Roy, R., Zakoan, J.-M., 2005. Diagnostic checking in arma models with uncorrelated errors. Journal of the American Statistical Association 100 (470), pp. 532–544.

Godfrey, L. G., 1979. Testing the adequacy of a time series model. Biometrika 66 (1), pp. 67–72.

Grenander, U., Rosenblatt, M., 1957. Statistical analysis of stationary time series. John Wiley and Sons, New York.

Hannan, E. J., 1969. The identification of vector mixed autoregressive-moving average systems. Biometrika 56 (1), pp. 223–225.

Hillmer, S. C., Tiao, G. C., 1979. Likelihood function of stationary multiple autoregressive moving average models. Journal of the American Statistical Association 74 (348), 652–660.

Hong, Y., 1996. Consistent testing for serial correlation of unknown form. Econometrica 64 (4), 837–837.

Hosking, J. R. M., 1980. The multivariate portmanteau statistic. Journal of the American Statistical Association 75 (371), pp. 602–608.

Hosking, J. R. M., 1981. Equivalent forms of the multivariate portmanteau statistic. Journal of the Royal Statistical Society. Series B (Methodological) 43 (2), pp. 261–262.

Kheoh, T. S., McLeod, A., 1992. Comparison of two modified portmanteau tests for model adequacy. Computational Statistics & Data Analysis 14 (1), 99 – 106.

Klein, A., Spreij, P., 2004. An explicit expression for the fisher information matrix of a multiple time series process. Discussion paper, Universiteit van Amsterdam, Faculty of Economics and Econometrics.

Kwan, A. C., Sim, A.-B., 1996. Portmanteau tests of randomness and jenkins' variance-stabilizing transformation. Economics Letters 50 (1), 41 – 49.

Kwan, A. C., Sim, A.-B., Wu, Y., 2005. A comparative study of the finite-sample performance of some portmanteau tests for randomness of a time series. Computational Statistics and Data Analysis 48 (2), 391 – 413.

Kwan, A. C., Wu, Y., 2003. A re-examination of the finite-sample properties of peña and rodríguez's portmanteau test of lack of fit for time series. Departmental working papers, Chinese University of Hong Kong, Department of Economics.

Kwan, A. C. C., Wu, Y., 1997. Further results on the finite-sample distribution of monti's portmanteau test for the adequacy of an *arma* (p,q) model. Biometrika 84 (3), pp. 733–736.

Lee, S., Ng, C. T., 2010. Trimmed portmanteau test for linear processes with infinite variance. Journal of Multivariate Analysis 101 (4), 984 – 998.

Li, W., 2004. Diagnostic checks in time series. Chapman and Hall.

Li, W. K., McLeod, A. I., 1981. Distribution of the residual autocorrelations in multivariate arma time series models. Journal of the Royal Statistical Society. Series B (Methodological) 43 (2), pp. 231–239.

Lin, J.-W., McLeod, A., 2006. Improved penarodriguez portmanteau test. Computational Statistics & Data Analysis 51 (3), 1731 – 1738.

Lin, J.-W., McLeod, A. I., 2008. Portmanteau tests for arma models with infinite variance. Journal of Time Series Analysis 29 (3), 600–617.

Ljung, G. M., 1986. Diagnostic testing of univariate time series models. Biometrika 73 (3), pp. 725–730.

Ljung, G. M., Box, G. E. P., 1978. On a measure of lack of fit in time series models. Biometrika 65 (2), pp. 297–303.

Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer.

Lütkepohl, H., 2006. Chapter 6 forecasting with varma models. Vol. 1 of Handbook of Economic Forecasting. Elsevier, pp. 287 – 325.

Mahdi, E., McLeod, A., 2012. Improved multivariate portmanteau test. Journal of Time Series Analysis 33 (2), 211–222.

Mahdi, E., McLeod, A., 2013. Package 'portes'. Date: Abril 17, 2013.

Mauricio, J. A., 1995. Exact maximum likelihood estimation of stationary vector arma models. Journal of the American Statistical Association 90 (429), pp. 282–291.

McLeod, A. I., 1979. Distribution of the residual cross-correlation in univariate arma time series models. Journal of the American Statistical Association 74 (368), pp. 849–855.

Milhøj, A., 1981. A test of fit in time series models. Biometrika 68 (1), pp. 177–187.

Monti, A. C., 1994. A proposal for a residual autocorrelation test in linear models. Biometrika 81 (4), pp. 776–780.

Newbold, P., 1980. The equivalence of two tests of time series model adequacy. Biometrika 67 (2), pp. 463–465.

Nicholls, D. F., Hall, A. D., 1979. The exact likelihood function of multivariate autoregressive-moving average models. Biometrika 66 (2), 259–264.

Paparoditis, E., 2005. Testing the fit of a vector autoregressive moving average model. Journal of Time Series Analysis 26 (4), 543 – 568.

Peña, D., Rodríguez, J., 2002. A powerful portmanteau test of lack of fit for time series. Journal of the American Statistical Association 97 (458), pp. 601–610.

Peña, D., Rodríguez, J., 2006. The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series. Journal of Statistical Planning and Inference 136 (8), 2706 – 2718.

Pierce, D. A., 1970. A duality between autoregressive and moving average processes concerning their least-squares parameter estimates. The Annals of Mathematical Statistics 41, pp. 722–726.

Poskitt, D. S., Tremayne, A. R., 1980. Testing the specification of a fitted autoregressive-moving average model. Biometrika 67 (2), 359–363.

Poskitt, D. S., Tremayne, A. R., 1982. Diagnostic tests for multiple time series models. The Annals of Statistics 10 (1), pp. 114–120.

Priestley, M., 1981. Spectral Analysis and Time Series. Academic Press: New York.

Reinsel, G., 1997. Elements of Multivariate Time Series Analysis, 2nd edn. Springer Verlag, New York.

Shorack, G., Wellner, J., 1986. Empirical Processes with Applications to Statistics. New York: Wiley.

Ubierna, A., Velilla, S., 2007. A goodness-of-fit process for arma models based on a modified residual autocorrelation sequence. Journal of Statistical Planning and Inference 137 (9), 2903 – 2919.

Velilla, S., 1994. A goodness-of-fit test for autoregressive moving-average models based on the standardized sample spectral distribution of the residuals. Journal of Time Series Analysis 15, 637–647.

Walker, A., 1952. Some properties of the asymptotic power functions of goodness-of-fit tests for linear autoregressive schemes. Journal of the Royal Statistical Society. Series B (Methodological) 14 (1), pp. 117–134.

Whittle, P., 1952. Tests of fit in time series. Biometrika 39 (3-4), 309–318.